FUNDAMENTAL LIMITS OF EXACT-REPAIR REGENERATING CODES

A Dissertation

by

SHUO SHAO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Tie Liu
Committee Members,	Xiaoning Qian
	Jean-Francois Chamberland-Tremblay
	Michael Anshelevich
Head of Department,	Miroslav Begovic

May 2017

Major Subject: Electrical Engineering

ABSTRACT

Understanding the fundamental limits of communication systems involves both constructing efficient coding schemes as well as proving mathematically that certain performance is impossible to achieve; the latter is known as the converse problem in information theory. This thesis focused on the converse problems for complex information systems such as self-repair distributed storage and coded caching systems, and our goal was to establish tight converse results for such systems by exploiting problem-specific combinatorial structures.

The main part of this thesis dealt with exact-repair regenerating codes, which were first proposed by Dimakis et al. in 2010. In particular, we considered two extensions of the original setting of Dimakis et al., namely 1) multilevel diversity coding with regeneration and 2) secure exact-repair regenerating codes. For the problem of multilevel diversity coding with regeneration, we showed, via the proposed combinatorial approach, that the natural separate encoding strategy can achieve the optimal tradeoff between the normalized storage capacity and repair bandwidth at the minimum-bandwidth rate (MBR) point. This settled a conjecture by Tian and Liu in 2015.

For the problem of secure exact-repair regenerating codes, all known results from the literature showed that the achievable tradeoff regions between the normalized storage capacity and repair bandwidth have a single corner point, achieved by a scheme proposed by Shah, Rashmi and Kumar (the SRK point). Since the achievable tradeoff regions of the exact-repair regenerating code problem without any secrecy constraints were known to have multiple corner points in general, these existing results suggested a phase-change-like behavior, i.e., enforcing a secrecy constraint immediately reduces the tradeoff region to one with a single corner point. In our work, we first showed that when the secrecy

parameter is sufficiently large, the SRK point is indeed the only corner point of the tradeoff region. However, when the secrecy parameter is small, we showed that the tradeoff region can, in fact, have multiple corner points. In particular, we established a precise characterization of the tradeoff region for a particular problem instance, which has exactly two corner points. Thus, a smooth transition, instead of a phase-change-type of transition, should be expected as the secrecy constraint is gradually strengthened.

DEDICATION

To my mother and my father. I give my best and most gratitude to you for your support.

ACKNOWLEDGMENTS

I would like to thank the ECE Department of Texas A&M University for all the help from them in my three years working and studying experience here. Special thanks to my supervisor, Dr Tie Liu, and my committee member, Dr Xiaoning Qian, Dr Jean-Francois Chamberland-Tremblay, Dr Michael Anshelevich. Last but not the least, I would like to thank my parents and my friends for their support.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Tie Liu and Professor Jean-Francois Chamberland-Tremblay and Professor Xiaoning Qian of the Department of Electric Engineering and Professor Michael Anshelevich of the Department of Math. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University and by the National Science Foundation under Grants CCF-13-20237 and CCF-15-24839.

NOMENCLATURE

$\mathbb{N}$	The set of natural numbers
MDC	Multilevel Diversity Coding
DS	Distributed Storage
MBR	Minimum Bandwidth Rate
MSR	Minimum Storage Rate
Inner Bound	Sufficient condition for an *achievable* rate
Outer Bound	Necessary condition for an *achievable* rate
$H(X)$	Entropy of X, computed as $$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$
$\binom{n}{k}$	Binomial number, the number of different ways of choosing a k-elements subset from a total of n elements
$[N]$	Set $\{1, 2, \ldots, N\}$ for $N \in \mathbb{N}$

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

1.1 Exact-Repair Regenerating Codes

The focus of this thesis is distributed and cloud storage systems. Fault tolerance and node repair are two fundamental ingredients of reliable distributed storage systems. While the study of fault tolerance via diversity coding has been in the literature for decades [1–6], systematic studies of node repair mechanisms were started only recently by Dimakis et al. in their pioneering work [7]. A particular model, which has received a significant amount of attention in the literature, is the so-called *exact-repair regenerating code* problem.

More specifically, in an (n, k, d) exact-repair regenerating code problem, a file M of size B is to be encoded and then stored in a total of n distributed storage nodes, each of capacity α. The encoding needs to ensure that: 1) the file M can be perfectly recovered by having *full* access to any k out of the total n storage nodes; 2) when a node failure occurs, the failed node can be regenerated by *extracting* data of size β from each of an arbitrary set of d remaining nodes. These two system requirements, the file recovery and node regeneration, are shown in Figure 1.1 respectively.

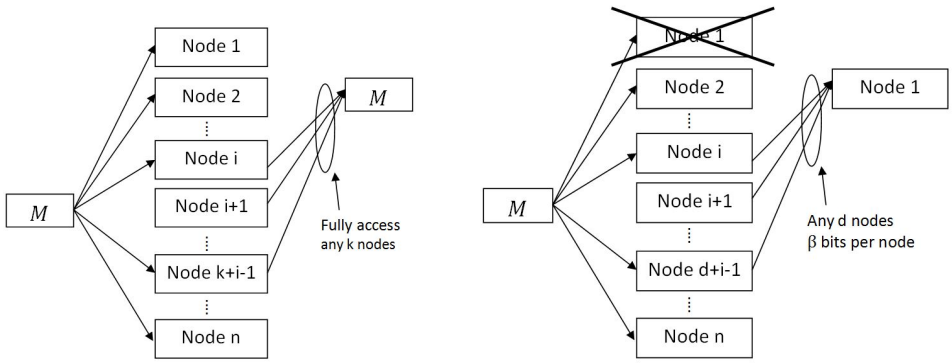


Figure 1.1: System requirement for file recovery function (left) and node regenerating function (right) for an (n, k, d) regenerating code.

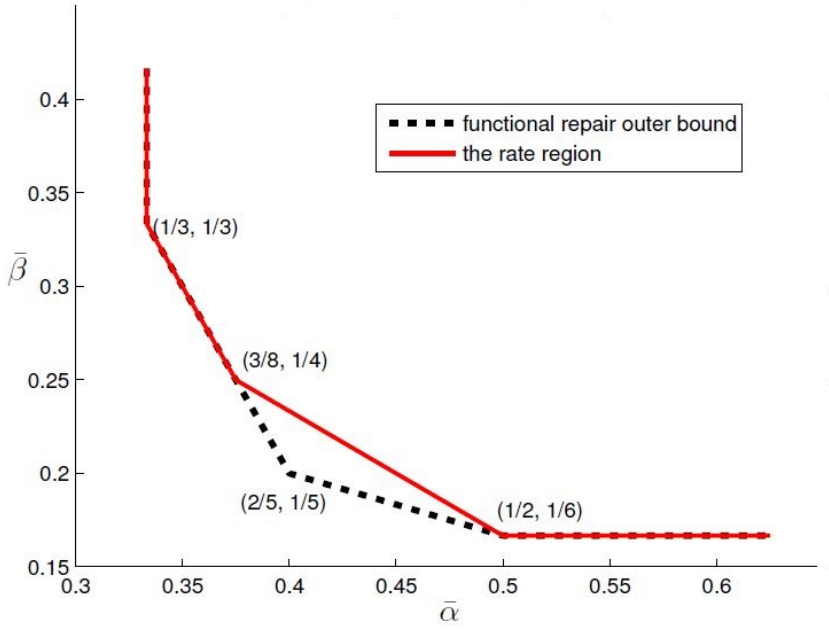


Figure 1.2: The regions above the solid line is the optimal normalized storage-capacity repair-bandwidth tradeoff region for the $(4, 3, 3)$ exact-repair regenerating code, while the dashed line is the outer bound obtained by *cut-set bound* approach. Reprinted with permission from [8].

Our goal is to characterize the region for all possible normalized *storage-capacity* and *bandwidth rate* pair defined as

$$\left(\bar{\alpha}, \bar{\beta}\right) = \left(\frac{\alpha}{B}, \frac{\beta}{B}\right) \tag{1.1}$$

where B is the size of the file. An important technical contribution of [7] was to show that there is an inherent *tradeoff* between the node capacity α and the repair bandwidth β in satisfying both the file-recovery and node-regeneration requirements. In particular, it has been shown [9] that the achievable normalized storage-capacity repair-bandwidth trade-off region for any (n, k, d) exact-repair regenerating code problem with $k > 1$ features *multiple* corner points including the the all-important *minimum storage rate (MSR)* and *minimum bandwidth rate (MBR)* points.

Fig. 1.2 illustrates the optimal normalized storage-capacity repair-bandwidth tradeoff region (the region above the solid line) for the $(4, 3, 3)$ exact-repair regenerating code problem, which features three corner points including the MSR point $(1/3, 1/3)$ and the MBR point $(1/2, 1/6)$. To prove that the optimal tradeoffs are indeed precisely characterized by the solid lines, one has to show that: 1) any rate pair $\left(\bar{\alpha}, \bar{\beta}\right)$ above the solid line is *achievable*. In particular, this can be accomplished by showing that all *corner points* of the rate region are achievable; 2) any rate pair $\left(\bar{\alpha}, \bar{\beta}\right)$ below the solid lines *cannot* be achieved by any coding scheme. In the information theory literature, this is usually known as the *converse* problem and is the main focus of this thesis.

In information theory, the standard approach for establishing converse results is via the so-called *cut-set* bounds [10]. The cut-set bounds are a set of *outer* bounds on the tradeoff region obtained via the graph-theoretic notion of *cuts*. The dotted lines in Fig. 1.2 illustrate the cut-set bounds for the $(4, 3, 3)$ exact-repair regenerating code problem. As we can see, the cut-set bounds, though readily available, are *strictly suboptimal* in terms of characterizing the optimal tradeoffs for the exact-repair regenerating code problem.[1] Instead, the optimal tradeoffs as illustrated by the solid lines were established by Tian [8] via a recently proposed *computational* approach.

In Tian's computational approach [8], the converse problem is formulated as linear programs (LPs). The constraints of the LPs are given by the system requirements written in terms of the (Shannon) entropies [10], as well as the *intrinsic* relations among the entropies. The nature of the computational approach is to develop efficient algorithms to the solve the LPs by exploring the built-in *symmetries* of the problem. Unfortunately, the size of the LPs grows extremely fast with the number of the nodes n in the system. Given today's computational capabilities, the proposed computational approach [8] does

[1] The cut-set bounds, however, are indeed optimal under a weaker notion of the repair requirement known as the *functional repair* [7].

not appear to be viable for $n \geq 10$.

We conclude this section by mentioning that despite intensive research efforts that have yielded many highly non-trivial partial results [7–9, 11], the optimal tradeoffs between the node capacity α and repair bandwidth β have *not* been fully understood for the general (n, k, d) exact-repair regenerating code problem.

1.2 Thesis Outline

In this thesis we shall focus on two *extensions* of the exact-repair regenerating code problem [7], namely *multilevel diversity coding with regeneration* and *secure exact-repair regenerating codes*. In particular, in Chapter 2 we shall address the optimality of a natural *superposition* coding scheme for the problem of multilevel diversity coding with regeneration. Our results have been presented at the 50th Annual Conference on Information Sciences and Systems (CISS) in 2016 [12]. In Chapter 3, we shall address an open conjecture in the literature [13, 14] regarding to a *polyhedral* property of the tradeoff region of the secure exact-repair regenerating code problem. Our results have been submitted to the 2017 IEEE International Symposium on Information Theory (ISIT) and the IEEE Transactions on Information Theory.

From the methodology viewpoint, our focus is on the converse problems which we view as *combinatorial* problems. Different from the traditional approach that relies on the cut-set bounds [7], our goal is to obtain *tight* converse results by exploring the *problem-specific* combinatorial structures of the problems. Also different from the recently proposed computational approach [8] which is numerical by nature, our approach is *analytical* and is targeted at problems of *general* parameters.

Finally in Chapter 4, we shall consider a related problem known as *coded caching* in the literature [15]. In particular, we shall use the combinatorial approach to address the optimality of the so-called Maddah-Ali-Niesen coding scheme [15] when there are a large

number of users in the system.

2. MULTILEVEL DIVERSITY CODING WITH REGENERATING

2.1 Introduction

The multilevel diversity regenerating code[1] problem was first introduced by Tian and Liu in [16] (where it was called *multilevel diversity coding with regeneration (MLDR)*). In an (n, d) multilevel diversity regenerating code problem, a total of d independent files $M_1, \ldots, M_d$ of size $B_1, \ldots, B_d$, respectively, are to be stored in n distributed storage nodes, each of capacity α. The encoding needs to ensure that the file M_k can be perfectly recovered by having full access to any k out of the total n storage nodes for any $k \in \{1, \ldots, d\}$. In addition, when node failures occur and there are only d remaining nodes in the system, it is required the data originally stored in any failed node can be recovered by downloading data of size β from each one of the d remaining nodes. Figure 2.1 illustrate the system requirement for file recovery and node regeneration respectively.

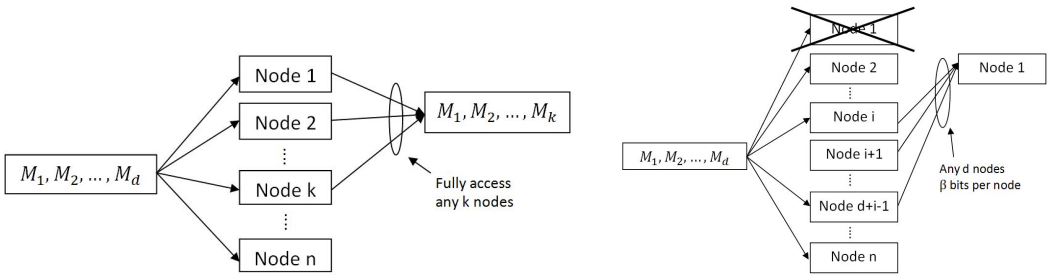


Figure 2.1: System requirement for file recovery (left) and node regenerating (right) for a multilevel diversity coding with regenerating system.

Based on the above description, it should be clear that an (n, k, d) regenerating code

[1]Reprinted with permission from "Multilevel diversity coding with regeneration: Separate coding achieves the MBR point," by S. Shao, T. Liu, and C. Tian, in *Proc. 50th Ann. Conf. Inf. Sci. Systems (CISS)*, Princeton, NJ, USA, Mar. 2016, pp. 602–607, copyright [2016] by *IEEE*.

problem can be thought of as a special case of the (n, d) multilevel diversity regenerating code problem with $B_k = B$ and $B_j = 0$ for all $j \neq k$. From the code construction perspective, it is thus natural to consider the so-called *separate* encoding scheme for the multilevel diversity regenerating code problem. That is, to construct an (n, d) multilevel diversity regenerating code, we may simply use an (n, k, d) regenerating code to encode file M_k for each $k \in \{1, \ldots, d\}$, and the coded messages for each file remain separate when stored in the storage nodes and during the repair processes.

2.2 Problem Formulation and Known Results

Let $(n, d, N_1, \ldots, N_d, T, S)$ be a tuple of positive integers such that $n \geq d + 1 \geq 3$. Formally, an $(n, d, N_1, \ldots, N_d, T, S)$ multilevel diversity regenerating code consists of:

- for each $i \in [1 : n]$, a message-encoding function $f_i : \prod_{k=1}^{d} [1 : N_k] \to [1 : T]$;
- for each $\mathcal{A} \subseteq [1 : n] : |\mathcal{A}| \in [1 : d]$, a message-decoding function $g_{\mathcal{A}} : [1 : T]^{|\mathcal{A}|} \to \prod_{k=1}^{|\mathcal{A}|} [1 : N_k]$;
- for each $\mathcal{B} \subseteq [1 : n] : |\mathcal{B}| = d$, $i \in \mathcal{B}$, and $j \in 1 : [n] \setminus \mathcal{B}$, a repair-encoding function $f_{i \to j}^{\mathcal{B}} : [1 : T] \to [1 : S]$; and
- for each $\mathcal{B} \subseteq [1 : n] : |\mathcal{B}| = d$ and $j \in [1 : n] \setminus \mathcal{B}$, a repair-decoding function $g_j^{\mathcal{B}} : [1 : S]^d \to [1 : T]$.

For each $k \in [1 : d]$, let M_k be a message that is uniformly distributed over $[1 : N_k]$. The messages $\mathsf{M}_1, \ldots, \mathsf{M}_d$ are assumed to be mutually independent. For each $i \in [1 : n]$, let $\mathsf{W}_i = f_i(\mathsf{M}_1, \ldots, \mathsf{M}_d)$ be the data stored at the ith storage node, and for each $\mathcal{B} \subseteq [1 : n] : |\mathcal{B}| = d$, $i \in \mathcal{B}$, and $j \in [1 : n] \setminus \mathcal{B}$, let $\mathsf{S}_{i \to j}^{\mathcal{B}} = f_{i \to j}^{\mathcal{B}}(\mathsf{W}_i)$ be the data downloaded from the ith storage node in order to regenerate the data originally stored at the jth storage node

under the context of repair group $\mathcal{B}$. Obviously,

$$(B_k = \log N_k : k \in [1:d]), \quad \alpha = \log T, \quad \text{and} \ \ \beta = \log S$$

represent the message rates, storage capacity, and repair bandwidth, respectively.

A normalized message-rate storage-capacity repair-bandwidth tuple $(\bar{B}_1, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta})$ is said to be *achievable* if an $(n, d, N_1, \ldots, N_d, T, S)$ multilevel diversity regenerating code can be found such that:

$$\bar{B}_k = \frac{B_k}{\sum_{i=1}^{d} B_i} \ \forall k \in [1:d], \tag{2.1}$$

$$\bar{\alpha} = \frac{\alpha}{\sum_{i=1}^{d} B_i}, \quad \bar{\beta} = \frac{\beta}{\sum_{i=1}^{d} B_i}, \tag{2.2}$$

$$(\mathsf{M}_1, \ldots, \mathsf{M}_{|\mathcal{A}|}) = g_{\mathcal{A}}(\mathsf{W}_i : i \in \mathcal{A})$$
$$\forall \mathcal{A} \subseteq [1:n] : |\mathcal{A}| \in [1:d], \quad \text{and} \tag{2.3}$$

$$\mathsf{W}_j = g_j^{\mathcal{B}}(\mathsf{S}_{i \to j}^{\mathcal{B}} : i \in \mathcal{B})$$
$$\forall \mathcal{B} \subseteq [1:n] : |\mathcal{B}| = d \ \text{ and } \ j \in [1:n] \setminus \mathcal{B}. \tag{2.4}$$

The closure of all achievable $(\bar{B}_1, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta})$ tuple is the achievable normalized message-rate storage-capacity repair-bandwidth tradeoff region $\mathcal{R}_{n,d}$ for the (n, d) multilevel diversity regenerating code problem. For a fixed normalized message-rate tuple $(\bar{B}_1, \ldots, \bar{B}_d)$, the achievable normalized storage-capacity repair-bandwidth tradeoff region is the collection of all normalized storage-capacity repair-bandwidth pairs $(\bar{\alpha}, \bar{\beta})$ such that $(\bar{B}_1, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,d}$ and is denoted by $\mathcal{R}_{n,d}(\bar{B}_1, \ldots, \bar{B}_d)$.

As mentioned previously in the introduction, an (n, k, d, N, T, S) regenerating code can be thought of as an $(n, d, N_1, \ldots, N_d, T, S)$ multilevel diversity regenerating code with $N_k = N$ and $N_i = 1$ for all $i \in [1:d] \setminus \{k\}$. Thus, for the (n, k, d) regenerat-

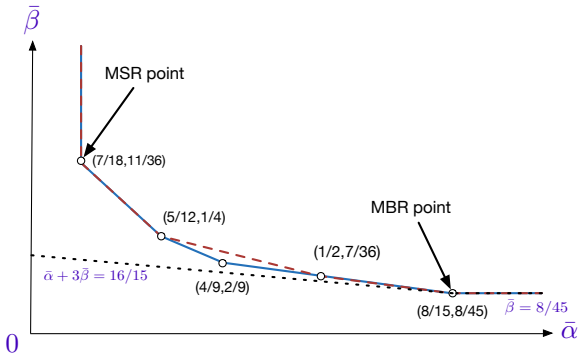


Figure 2.2: The optimal tradeoff curve between the normalized storage-capacity $\bar{\alpha}$ and repair-bandwidth $\bar{\beta}$ (the solid line) and the best possible tradeoffs that can be achieved by separate coding (dashed line) for the $(4, 3)$ multilevel diversity coding problem with $(\bar{B}_1, \bar{B}_2, \bar{B}_3) = (0, 1/3, 2/3)$ [16]. The two new outer bounds (2.9) and (2.10) intersect precisely at the MBR point $(8/15, 8/45)$. Graph is cited from [12].

ing code problem, the achievable normalized storage-capacity repair-bandwidth tradeoff region $\mathcal{R}_{n,k,d}$ is simply given by $\mathcal{R}_{n,d}(0, \ldots, 0, \bar{B}_k = 1, 0, \ldots, 0)$.

Based on the connection between the regenerating code and the multilevel diversity regenerating code problems mentioned above, a natural approach for constructing an $(n, d, N_1, \ldots, N_d, T, S)$ multilevel diversity regenerating code is to use to an (n, k, d, N_k, T_k, S_k) regenerating code to encode the message M_k separately for each $k \in [1 : d]$. Since the coded data are kept separate during the encoding and repair processes, we have

$$T = \prod_{k=1}^{d} T_k \quad \text{and} \quad S = \prod_{k=1}^{d} S_k.$$

Thus, for the (n, d) multilevel diversity regenerating code problem, the separate coding normalized storage-capacity repair-bandwidth tradeoff region $\hat{\mathcal{R}}_{n,d}(\bar{B}_1, \ldots, \bar{B}_d)$ for

a fixed normalized message-rate tuple $(\bar{B}_1, \ldots, \bar{B}_d)$ is given by:

$$\left(\left(\sum_{k=1}^{d} \bar{\alpha}_k \bar{B}_k, \sum_{k=1}^{d} \bar{\beta}_k \bar{B}_k \right) : (\bar{\alpha}_k, \bar{\beta}_k) \in \mathcal{R}_{n,k,d} \right). \tag{2.5}$$

Despite being a natural scheme, it was shown in [16] that separate coding is in general *suboptimal* in achieving the optimal tradeoffs between the normalized storage-capacity and repair-bandwidth. Figure 2.2 shows the optimal tradeoff curve between the normalized storage-capacity and repair-bandwidth and the best possible tradeoffs that can be achieved by separate coding for $(4, 3)$ multilevel diversity regenerating code problem with $(\bar{B}_1, \bar{B}_2, \bar{B}_3) = (0, 1/3, 2/3)$ [16]. Clearly, for this example, separate coding is strictly suboptimal when $\bar{\alpha} \in (5/12, 1/2)$. On the other hand, when $\bar{\alpha} \leq 5/12$ or $\bar{\alpha} \geq 1/2$, separate coding can in fact achieve the optimal tradeoffs. In particular, for this example, separate encoding achieves the minimum-storage-regenerating (MSR) point $(7/18, 11/36)$ and the minimum-bandwidth-regenerating (MBR) point $(8/15, 8/45)$.

That separate coding can achieve the MSR point for the above example is, in fact, not incidental. For a general (n, d) multilevel diversity regenerating code problem, it was shown in [16] that any achievable normalized message-rate storage-capacity repair-bandwidth tuple $(\bar{B}_1, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,d}$ must satisfy:

$$\bar{\alpha} \geq \sum_{k=1}^{d} \frac{\bar{B}_k}{k} \tag{2.6}$$

$$\text{and} \quad (d-1)\bar{\alpha} + \bar{\beta} \geq \sum_{k=1}^{d} \frac{(d-1)(d+1-k)+1}{k(d+1-k)} \bar{B}_k. \tag{2.7}$$

When set as equalities, the intersection of (2.6) and (2.7) is given by

$$(\bar{\alpha}, \bar{\beta}) = \left(\sum_{k=1}^{d} \frac{\bar{B}_k}{k}, \sum_{k=1}^{d} \frac{(d+1-k)\bar{B}_k}{k} \right).$$

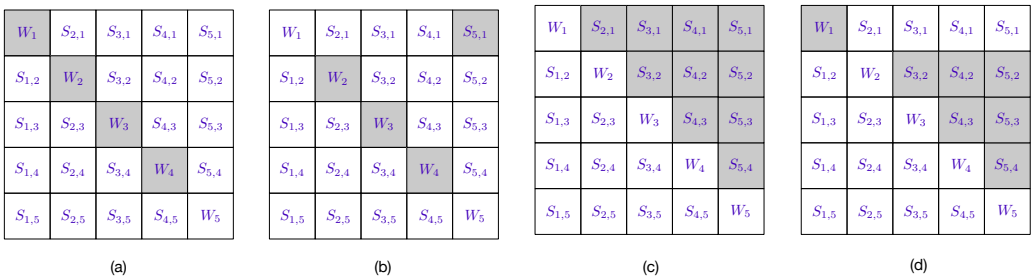


Figure 2.3: The repair diagram of Duursma [11] for the $(5, 4)$ multilevel diversity regenerating code problem. The collections of coded data in (2.12)–(2.15) for $k = 4$ are illustrated in (a)–(d), respectively.

For any $k \in [1 : d]$, the MSR point for the (n, k, d) regenerating code problem can be written as [7]:

$$\left(\frac{1}{k}, \frac{d+1-k}{k}\right) \in \mathcal{R}_{n,k,d}. \tag{2.8}$$

We may thus conclude immediately from (2.5) that separate coding can achieve the MSR point for the general (n, d) multilevel diversity regenerating code problem.

As mentioned previously in the introduction, an (n, k, d, N, T, S) regenerating code can be thought of as an $(n, d, N_1, \ldots, N_d, T, S)$ multilevel diversity regenerating code with $N_k = N$ and $N_i = 1$ for all $i \in [1 : d] \setminus \{k\}$. Thus, for the (n, k, d) regenerating code problem, the achievable normalized storage-capacity repair-bandwidth tradeoff region $\mathcal{R}_{n,k,d}$ is simply given by $\mathcal{R}_{n,d}(0, \ldots, 0, \bar{B}_k = 1, 0, \ldots, 0)$.

2.3 New Results

Our main result for this section is to show that separate coding can achieve the MBR point for a general (n, d) multilevel diversity regenerating code problem as well. From the technical viewpoint, this is mainly accomplished by establishing the following two new outer bounds for the achievable normalized message-rate storage-capacity repair-

bandwidth tradeoff region $\mathcal{R}_{n,d}$.

Theorem 1. *For a general (n, d) multilevel diversity regenerating code problem, any achievable normalized message-rate storage-capacity repair-bandwidth tuple $(\bar{B}_1, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,d}$ must satisfy:*

$$\bar{\beta} \geq \sum_{k=1}^{d} T_{d,k}^{-1} \bar{B}_k, \tag{2.9}$$

$$\textit{and} \quad \bar{\alpha} + J_{d-1}\bar{\beta} \geq J_d \sum_{k=1}^{d} T_{d,k}^{-1} \bar{B}_k \tag{2.10}$$

where $J_d := \sum_{i=1}^{d} i$ and $T_{d,k} := \sum_{i=1}^{k}(d+1-i)$.

When set as equalities, the intersection of (2.9) and (2.10) is given by:

$$\left(\bar{\alpha}, \bar{\beta}\right) = \left(d\sum_{k=1}^{d} T_{d,k}^{-1} \bar{B}_k, \sum_{k=1}^{d} T_{d,k}^{-1} \bar{B}_k \right).$$

For any $k \in [d]$, the MBR point for the (n, k, d) regenerating code problem can be written as [7]

$$\left(dT_{d,k}^{-1}, T_{d,k}^{-1}\right) \in \mathcal{R}_{n,k,d}. \tag{2.11}$$

We may thus conclude immediately that separate coding can achieve the MBR point for the general (n, d) multilevel diversity regenerating code problem.

For the example considered in Figure 2.2, the outer bounds (2.9) and (2.10) have also been plotted in the same figure. As illustrated, they intersect precisely at the MBR point $(8/15, 8/45)$. Interestingly, for this example at least, the outer bound (2.10) is tight *only* at the MBR point.

Our proof of the theorem is rather long and technical, meanwhile some lemmas are

applied to proceed the proof of our theorem. All proof of the lemmas is deferred to Appendix A to enhance the flow of the paper. The main ingredients of the proof are summarized below:

1) First note that the outer bounds (2.9) and (2.10) are independent of the total number of storage nodes n in the system. In fact, in our proof, we only need to focus on the cases where $n = d + 1$. For the cases where $n > d + 1$, since any subsystem consisting of $d+1$ out of the total n storage nodes must give rise to an $(n = d+1, d)$ multilevel diversity regenerating code problem, the outer bounds (2.9) and (2.10) must apply as well.

2) When $n = d + 1$, any repair group $\mathcal{B}$ of size d is uniquely determined by the node j to be repaired, i.e., $\mathcal{B} = [1 : n] \setminus \{j\}$, and hence can be dropped from the notation $\mathsf{S}_{i,j}^{\mathcal{B}}$ without causing any confusion. Furthermore, due to the built-in symmetry in the problem, to prove (2.9) and (2.10) we only need to consider the so-called *symmetrical* codes [8], for which the joint entropy of any subset of random variables from $\{\mathsf{M}_k : k \in [1 : d]\} \cup \{\mathsf{W}_i : i \in [1 : d + 1]\} \cup \{\mathsf{S}_{i,j} : i, j \in [1 : d+1], i \neq j\}$ remains *unchanged* under any permutation over the storage-node indices.

3) Our proof of the outer bounds (2.9) and (2.10) uses the classical *peeling* arguments, which are the landmarks of the converse proofs for multilevel diversity coding problems. The peeling argument was first introduced by Roche et al. [3] for the (symmetrical) multilevel diversity coding problem without any repair requirement, and was subsequently used in [16] to prove the outer bounds (2.6) and (2.7) for the multilevel diversity regenerating code problem. As a matter of fact, in [16] the outer bound (2.6) was proved by ignoring the repair requirement (2.4), and hence the result followed directly from [3, Theorem].

4) To apply the peeling arguments, one need to identify appropriate collections of coded data that allow *successive* recovery of the messages $(\mathsf{M}_1, \ldots, \mathsf{M}_d)$. To prove the outer bounds (2.6) and (2.7), the collections of coded data utilized in [3] and [16] were

$$\{\mathsf{W}_i : i \in [1:k]\} \tag{2.12}$$

and

$$\{\mathsf{W}_i : i \in [2:k]\} \cup \{\mathsf{S}_{i,1} : i \in [k+1:d+1]\} \tag{2.13}$$

for $k \in [1:d]$. To prove the new outer bounds (2.9) and (2.10), we shall consider the collections of coded data

$$\{\mathsf{S}_{i,j} : j \in [1:k], i \in [j+1:d+1]\} \tag{2.14}$$

and

$$\{\mathsf{W}_1\} \cup \{\mathsf{S}_{i,j} : j \in [2:k], i \in [j+1:d+1]\} \tag{2.15}$$

for $k \in [1:d]$. The structure of all four collections of coded data above can be nicely illustrated using the repair diagram introduced by Duursma [11]; see Figure 2.3. It is straightforward to verify that for each $k \in [1:d]$, the coded data $\{\mathsf{W}_i : i \in [1:k]\}$, and hence the message M_k, can be recovered from each one of the collections of coded data above.

2.4 Proof of Theorem 1

Proof of (2.9). To prove the outer bound (2.9), we shall apply a peeling argument that

utilizes the collection of coded data (2.14), which can now be compactly written as $\mathsf{L}_0^{(k)}$ using the newly introduced notations from above. The following telescoping result on $\mathsf{L}_0^{(k)}$ plays an essential role in our proof.

Proposition 1 (Telescoping over $\mathsf{L}_0^{(k)}$)**.** *For any symmetrical* $(n = d+1, d, (N_1, \ldots, N_d), T, S)$ *multilevel diversity regenerating code that satisfies the repair requirement* (2.4)*, we have*

$$T_{d,k}^{-1} H(\mathsf{L}_0^{(k)} | \mathsf{M}^{(k)}) \geq T_{d,k+1}^{-1} H(\mathsf{L}_0^{(k+1)} | \mathsf{M}^{(k)}) \tag{2.16}$$

for any $k \in [1 : d-1]$.

We now prove the following inequality by induction:

$$\beta \geq \sum_{j=1}^{k} T_{d,j}^{-1} B_j + T_{d,k}^{-1} H(\mathsf{L}_0^{(k)} | \mathsf{M}^{(k)}) \tag{2.17}$$

for any $k \in [1 : d]$. Note that

$$\begin{aligned} \beta &\overset{(a)}{\geq} T_{d,1}^{-1} \sum_{i=2}^{d+1} H(\mathsf{S}_{i,1}) \overset{(b)}{\geq} T_{d,1}^{-1} H(\mathsf{L}_1) \\ &\overset{(c)}{=} T_{d,1}^{-1} H(\mathsf{L}_1, \mathsf{M}_1) \overset{(d)}{=} T_{d,1}^{-1} B_1 + T_{d,1}^{-1} H(\mathsf{L}_1 | \mathsf{M}_1), \end{aligned}$$

and thus (2.17) holds for $k = 1$. Here, (a) follows from the repair-bandwidth constraints $H(\mathsf{S}_{i,1}) \leq \beta$ for $i \in [2 : d+1]$ and the fact that $T_{d,1} = d$; (b) is due to the union bound on entropy; (c) follows from the fact that M_1 is a function of W_1 and hence a function of L_1; and (d) is due to the chain rule for entropy and the fact that $H(\mathsf{M}_1) = B_1$.

Now assume that (2.17) holds for some $k \in [1 : d-1]$. Substituting the telescoping

result (2.16) into (2.17), we have

$$\begin{aligned}
\beta &\geq \sum_{j=1}^{k} T_{d,j} B_j + T_{d,k+1} H(\mathsf{L}_0^{(k+1)} | \mathsf{M}^{(k)}) \\
&\overset{(a)}{=} \sum_{j=1}^{k} T_{d,j} B_j + T_{d,k+1} H(\mathsf{L}_0^{(k+1)}, \mathsf{M}_{k+1} | \mathsf{M}^{(k)}) \\
&\overset{(b)}{=} \sum_{j=1}^{k} T_{d,j} B_j + T_{d,k+1} H(\mathsf{M}_{k+1} | \mathsf{M}^{(k)}) + \\
&\quad T_{d,k+1} H(\mathsf{L}_0^{(k+1)} | \mathsf{M}^{(k)}, \mathsf{M}_{k+1}) \\
&\overset{(c)}{=} \sum_{j=1}^{k+1} T_{d,j} B_j + T_{d,k+1} H(\mathsf{L}_0^{(k+1)} | \mathsf{M}^{(k+1)}),
\end{aligned}$$

which completes the induction and hence the proof of (2.17). Here, (a) follows from the fact that M_{k+1} is a function of $\mathsf{W}^{(k+1)}$, which is turn a function of $\mathsf{L}_0^{(k+1)}$; (b) is by the chain rule for conditional entropy; and (c) follows from the facts that all messages are independent and that $H(\mathsf{M}_{k+1}) = B_{k+1}$. This completes the induction step and hence the proof of (2.17).

Setting $k = d$ in (2.17) and by the fact that $H(\mathsf{L}_0^{(d)} | \mathsf{M}^{(d)}) \geq 0$, we have

$$\beta \geq \sum_{j=1}^{d} T_{d,j} B_j. \tag{2.18}$$

Normalizing both sides of (2.18) by $\sum_{k=1}^{d} B_k$ completes the proof of the outer bound (2.9).

Proof of (2.10). To prove the outer bound (2.10), we shall apply a peeling argument that utilizes the collections of coded data (2.14) and (2.15), which can now be compactly written as $\mathsf{L}_0^{(k)}$ and $\mathsf{L}_1^{(k)}$, respectively. In addition to Proposition 1, our proof also relies on the following telescoping result on $\mathsf{L}_1^{(k)}$.

Proposition 2 (Telescoping over $\mathsf{L}_1^{(k)}$)**.** *For any symmetrical* $(n = d{+}1, d, (N_1, \ldots, N_d), T, S)$

multilevel diversity regenerating code that satisfies the repair requirement (2.4)*, we have*

$$H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + (d-k)T_{d,k}^{-1}H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) \geq H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) \tag{2.19}$$

for any $k \in [1 : d-1]$.

We now prove the following inequality by induction:

$$\begin{aligned}\alpha + J_{d-1}\beta \geq J_d \sum_{j=1}^{k} T_{d,j}B_j + H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + \\ J_{d-k}T_{d,k}H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)})\end{aligned} \tag{2.20}$$

for any $k \in [1 : d]$. Note that

$$\begin{aligned}&\alpha + J_{d-1}\beta \\ &\overset{(a)}{\geq} H(\mathsf{W}_1) + J_{d-1}d^{-1}\sum_{i=2}^{d+1} H(\mathsf{S}_{i,1}) \\ &\overset{(b)}{\geq} H(\mathsf{W}_1) + J_{d-1}T_{d,1}^{-1}H(\mathsf{L}_1) \\ &\overset{(c)}{=} H(\mathsf{W}_1, \mathsf{M}_1) + J_{d-1}T_{d,1}^{-1}H(\mathsf{L}_1, \mathsf{M}_1) \\ &\overset{(d)}{=} H(\mathsf{M}_1) + H(\mathsf{W}_1|M_1) + J_{d-1}T_{d,1}^{-1}H(\mathsf{M}_1) + \\ &\quad J_{d-1}T_{d,1}^{-1}H(\mathsf{L}_1|\mathsf{M}_1) \\ &\overset{(e)}{=} (1 + J_{d-1}T_{d,1}^{-1})B_1 + H(\mathsf{W}_1|\mathsf{M}_1) + J_{d-1}T_{d,1}^{-1}H(\mathsf{L}_1|\mathsf{M}_1) \\ &\overset{(f)}{=} J_dT_{d,1}^{-1}B_1 + H(\mathsf{W}_1|\mathsf{M}_1) + J_{d-1}T_{d,1}^{-1}H(\mathsf{L}_1|\mathsf{M}_1),\end{aligned}$$

and thus (2.20) holds for $k = 1$. Here, (a) follows from the storage-capacity constraint $H(\mathsf{W}_1) \leq \alpha$ and the repair-bandwidth constraints $H(\mathsf{S}_{i,1}) \leq \beta$ for $i \in [2 : d+1]$; (b) is by the union bound on entropy and the fact that $T_{d,1} = d$; (c) follows from the fact that M_1

is a function of W_1 and hence a function of L_1; (d) is due to the chain rule for entropy; (d) is by the fact that $H(\mathsf{M}_1) = B_1$; and (f) follows from the fact that

$$\begin{aligned} 1 + J_{d-1}T_{d,1}^{-1} &= (T_{d,1} + J_{d-1})T_{d,1}^{-1} \\ &= (d + J_{d-1})T_{d,1}^{-1} = J_d T_{d,1}^{-1}. \end{aligned}$$

Now assume that (2.20) holds for some $k \in [1 : d-1]$. Substituting the telescoping result (2.19) into (2.20), we have

$$\begin{aligned} \alpha + J_{d-1}\beta \geq\ & J_d \sum_{j=1}^{k} T_{d,j}^{-1} B_j + H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + \\ & [J_{d-k} - (d-k)]\, T_{d,k}^{-1} H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) \\ =\ & J_d \sum_{j=1}^{k} T_{d,j}^{-1} B_j + H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + \\ & J_{d-1-k} T_{d,k}^{-1} H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}). \end{aligned} \tag{2.21}$$

Further substituting (2.16) into (2.21), we have

$$
\begin{aligned}
&\alpha + J_{d-1}\beta \\
&\geq J_d \sum_{j=1}^{k} T_{d,j}^{-1} B_j + H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + \\
&\quad J_{d-1-k} T_{d,k+1}^{-1} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) \\
&\overset{(a)}{=} J_d \sum_{j=1}^{k} T_{d,j}^{-1} B_j + H(\mathsf{L}_1^{(k+1)}, \mathsf{M}_{k+1}|\mathsf{M}^{(k)}) + \\
&\quad J_{d-1-k} T_{d,k+1}^{-1} H(\mathsf{L}_0^{(k+1)}, \mathsf{M}_{k+1}|\mathsf{M}^{(k)}) \\
&\overset{(b)}{=} J_d \sum_{j=1}^{k} T_{d,j}^{-1} B_j + (1 + J_{d-1-k} T_{d,k+1}^{-1}) H(\mathsf{M}_{k+1}|\mathsf{M}^{(k)}) + \\
&H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k+1)}) + J_{d-1-k} T_{d,k+1}^{-1} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k+1)}) \\
&\overset{(c)}{=} J_d \sum_{j=1}^{k} T_{d,j}^{-1} B_j + J_d T_{d,k+1}^{-1} H(\mathsf{M}_{k+1}|\mathsf{M}^{(k)}) + \\
&H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k+1)}) + J_{d-1-k} T_{d,k+1}^{-1} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k+1)}) \\
&\overset{(d)}{=} J_d \sum_{j=1}^{k+1} T_{d,j}^{-1} B_j + H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k+1)}) + \\
&\quad J_{d-1-k} T_{d,k+1}^{-1} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k+1)}),
\end{aligned}
$$

which completes the induction and hence the proof of (2.20). Here, (a) follows from the fact that M_{k+1} is a function of $\mathsf{W}^{(k+1)}$, which is in turn a function of $\mathsf{L}_1^{(k+1)}$ and further a function of $\mathsf{L}_0^{(k+1)}$; (b) is due to the chain rule for conditional entropy; (c) follows from the

fact that

$$\begin{aligned}1 + J_{d-1-k}T_{d,k+1}^{-1} &= (T_{d,k+1} + J_{d-1-k})T_{d,k+1}^{-1} \\ &= \left(\sum_{i=1}^{k+1}(d+1-i) + J_{d-1-k}\right)T_{d,k+1}^{-1} \\ &= \left(\sum_{i=d-k}^{d} i + J_{d-1-k}\right)T_{n,k+1}^{-1} = J_d T_{d,k+1}^{-1},\end{aligned}$$

and (d) is due to the facts that all messages are independent and that $H(\mathsf{M}_{k+1}) = B_{k+1}$. This completes the induction step and hence the proof of (2.20).

Set $k = d$ in (2.20). By the fact that $H(\mathsf{L}_1^{(d)}|\mathsf{M}^{(d)}) \geq 0$ and $J_0 = 0$, we have

$$\alpha + J_{d-1}\beta \geq J_d \sum_{j=1}^{k} T_{d,j}^{-1} B_j. \tag{2.22}$$

Normalizing both sides of (2.22) by $\sum_{k=1}^{d} B_k$ completes the proof of the outer bound (2.10).

We have thus completed the proof of Theorem 1.

3. SECURE EXACT-REPAIR REGENERATING CODES

3.1 Introduction

In this paper, we consider an extension of the aforementioned exact-repair regenerating code problem, which further requires certain security guarantee during the node-regeneration processes. The (n, k, d, ℓ) *secure exact-repair regenerating code* problem that we consider is the standard (n, k, d) exact-repair regenerating code problem [7–9, 11], with the additional constraint that the file M needs to be kept *information-theoretically* secure against an eavesdropper that can access the data extracted to regenerate a total of ℓ different failed nodes (possibly under different repair groups). Apparently, this is only possible when $\ell < k$. Furthermore, when $\ell = 0$, the secrecy constraint degenerates, and the (n, k, d, ℓ) secure exact-repair regenerating code problem reduces to the (n, k, d) exact-repair regenerating code problem without any security constraints.

More specifically, the (n, k, d, ℓ) *secure exact-repair regenerating code* problem has following constraints:

- file recovery: in this case, we introduce a random key K, which is applied to encipher the file M. Instead of storing file M directly into n nodes, we store a enciphered file. As shown in Figure 3.1, when we can fully access to any k nodes, we can recovery the deciphered file M again, but we don't have to recovery the random key.

- node regeneration: similar to exact-repair regenerating codes without secrecy constraint, there's a constraint on node regeneration. As shown in Figure 3.2, when any node is failed and we can extract data from any d other nodes for β bits per node at most, we can regenerate the failed node.

- information theoretic secure: the eavesdropper that can access the data extracted to

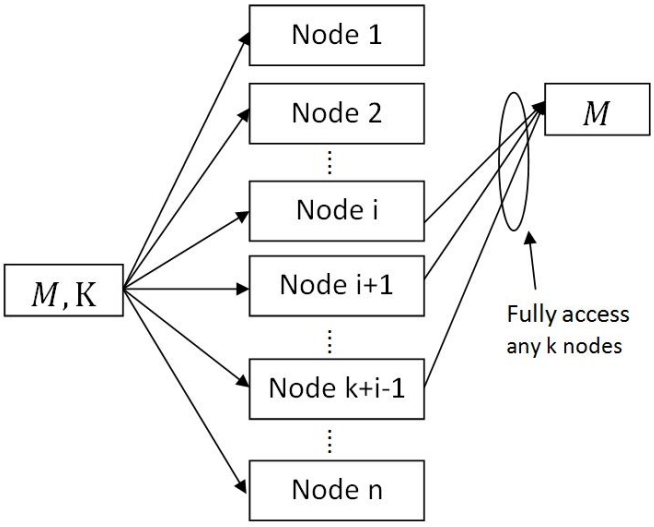


Figure 3.1: The file recovery constraint for (n, k, d) secure exact-repair regenerating codes.

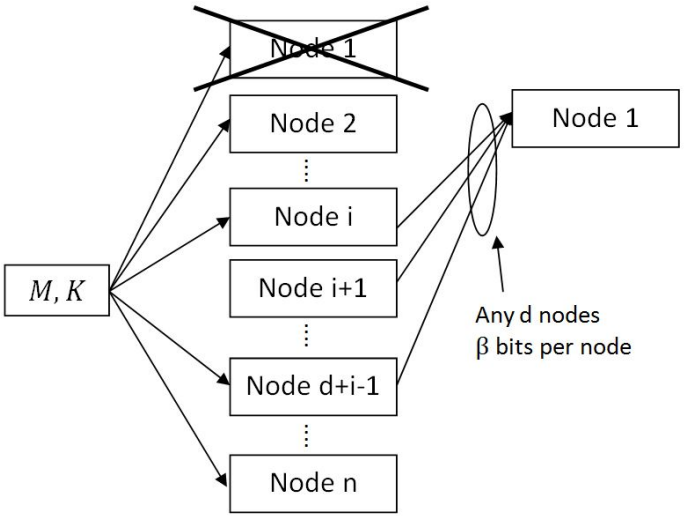


Figure 3.2: The node regeneration constraint for (n, k, d) secure exact-repair regenerating codes.

regenerate a total of ℓ different failed nodes. We require the mutual information between file M and collected data is zero for any ℓ nodes.

Under the additional secrecy constraint ($\ell \geq 1$), the optimal tradeoffs between the node capacity α and repair bandwidth β have been studied in [13, 14, 17–19]. In particular, Shah, Rashmi and Kumar [19] showed that a particular tradeoff point (referred to as the *SRK* point) can be obtained by extending an MBR code based on the product-matrix construction proposed in [9]. Later, it was shown that the SRK point is the *only* corner point of the tradeoff region for the cases where we have either $d = 2, 3$ [13], or $d = 4$ [14], or $k = 2$ [13], or $\ell = k - 1 = d - 1$ [13]. This is in sharp contrast to the original exact-repair regenerating code problem [7–9, 11] without any secrecy constraints, for which, as mentioned previously, the tradeoff region features *multiple* corner points when $k > 1$. Fig. 3.3 also illustrates the tradeoff region for the $(4, 3, 3, 1)$ secure exact-repair regenerating code problem, which features a single corner point at $(1, 1/3)$. Thus, the existing results from [13,14] seem to suggest a *phase-change-like* behavior that enforcing a secrecy constraint immediately reduces the tradeoff region from one with multiple corner points ($\ell = 0$) to one with a single corner point ($\ell \geq 1$).

The main results of this paper are two-folded.

- We first show, via new converse results, that for any given (k, d) pair, there is a lower bound on ℓ, denoted by $\ell^*(k, d)$, such that when $\ell \geq \ell^*(k, d)$, the SRK point is indeed the *only* conner point of the tradeoff region for the (n, k, d, ℓ) secure exact-repair regenerating code problem. As we shall see, the lower bound $\ell^*(k, d) \leq k-1$ for any (k, d) pair, and thus the tradeoff region for any (n, k, d, ℓ) problem with $\ell = k - 1$ or $k = 2$ must have a single corner point. In addition, the lower bound $\ell^*(k, d) = 1$ for any $d \in [2 : 4]$. Therefore, our result includes all previous results from [13] and [14] as special cases.

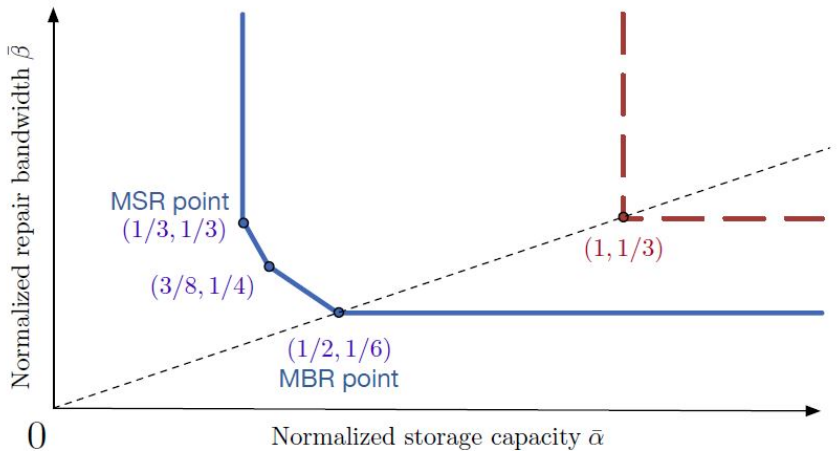


Figure 3.3: The regions above the solid and the dashed lines are the achievable normalized storage-capacity repair-bandwidth tradeoff regions for the $(4, 3, 3)$ exact-repair regenerating code without and with secrecy constraints respectively.

- Next, we show that when $1 \leq \ell < \ell^*(k, d)$, it is entirely possible that the tradeoff region features *multiple* corner points. In particular, we establish a precise characterization of the tradeoff region for the $(7, 6, 6, 1)$ problem, which has exactly *two* corner points (see Fig. 3.4 for an illustration). This result requires new achievability results as well as new converse results, the former of which are obtained by extending the layered coding scheme proposed in [20]. From the viewpoint of the rate region, our result suggests that a *smooth* transition, instead of a phase-change-type of transition, should be expected as the secrecy constraint is gradually strengthened by increasing the parameter ℓ.

3.2 Problem Formulation and Known Results

Let (n, k, d, N, K, T, S) be a tuple of positive integers such that $n \geq d+1 \geq k+1 \geq 2$. Formally, an (n, k, d, N, K, T, S) code consists of:

- for each $i \in [1 : n]$, a *message-encoding* function $f_i : [1 : N] \times [1 : K] \to [1 : T]$;

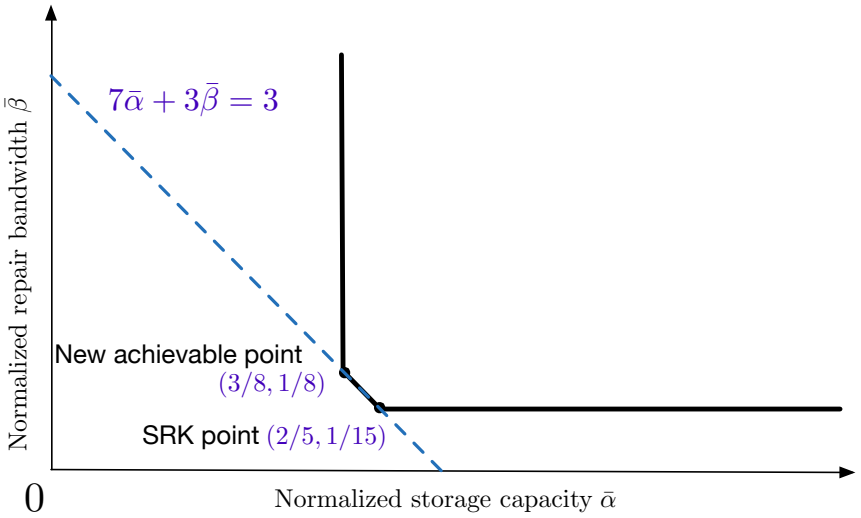


Figure 3.4: The regions above the solid line is the achievable normalized storage-capacity repair-bandwidth tradeoff region for the $(7, 6, 6, 1)$ secure exact-repair regenerating code problem. In addition to the SRK point $(2/15, 1/5)$, the tradeoff region has another corner point at $(3/8, 1/8)$.

- for each $\mathcal{A} \subseteq [1 : n]$ such that $|\mathcal{A}| = k$, a *message-decoding* function $g_{\mathcal{A}} : [1 : T]^k \to [1 : N]$;
- for each $\mathcal{B} \subseteq [1 : n]$ such that $|\mathcal{B}| = d$, $i \in \mathcal{B}$, and $j \in [1 : n] \setminus \mathcal{B}$, a *repair-encoding* function $f^{\mathcal{B}}_{i\to j} : [1 : T] \to [1 : S]$;
- for each $\mathcal{B} \subseteq [1 : n]$ such that $|\mathcal{B}| = d$ and $j \in [1 : n] \setminus \mathcal{B}$, a *repair-decoding* function $g^{\mathcal{B}}_j : [1 : S]^d \to [1 : T]$.

Let M be a message that is uniformly distributed over $[1 : N]$, and K be a secret key that is uniformly distributed over $[1 : K]$. The message M and the secret key K are assumed to be independent of each other. For each $i \in [1 : n]$, let $\mathsf{W}_i = f_i(\mathsf{M}, \mathsf{K})$ be the data stored at the ith storage node, and for each $\mathcal{B} \subseteq [1 : n]$ such that $|\mathcal{B}| = d$, $i \in \mathcal{B}$, and $j \in [1 : n] \setminus \mathcal{B}$, let $\mathsf{S}^{\mathcal{B}}_{i\to j} = f^{\mathcal{B}}_{i\to j}(\mathsf{W}_i)$ be the data extracted from the ith storage node in order to regenerate

the data stored at the jth storage node under the context of repair group $\mathcal{B}$. Obviously,

$$B = \log N, \quad \alpha = \log T, \quad \beta = \log S$$

represent the message rate, storage capacity, and repair bandwidth, respectively.

A normalized storage-capacity repair-bandwidth pair $(\bar{\alpha}, \bar{\beta})$ is said to be *achievable* for the (n, k, d, ℓ) secure exact-repair regenerating code problem if an (n, k, d, N, K, T, S) code can be found such that:

- (rate normalization)

$$\alpha/B = \bar{\alpha} \quad \text{and} \quad \beta/B = \bar{\beta}; \tag{3.1}$$

- (message recovery)

$$\mathsf{M} = g_{\mathcal{A}}(\mathsf{W}_i : i \in \mathcal{A}) \tag{3.2}$$

  for any $\mathcal{A} \subseteq [1 : n]$ such that $|\mathcal{A}| = k$;

- (node regeneration)

$$\mathsf{W}_j = g_j^{\mathcal{B}}(\mathsf{S}_{i\to j}^{\mathcal{B}} : i \in \mathcal{B}) \tag{3.3}$$

  for any $\mathcal{B} \subseteq [1 : n]$ such that $|\mathcal{B}| = d$ and $j \in [1 : n] \setminus \mathcal{B}$;

- (repair secrecy)

$$I(\mathsf{M}; (\mathsf{S}_{\to j} : j \in \mathcal{E})) = 0 \tag{3.4}$$

for any $\mathcal{E} \subseteq [1:n]$ such that $|\mathcal{E}| = \ell$, where $\mathsf{S}_{\to j} := (\mathsf{S}_{i\to j}^{\mathcal{B}} : \mathcal{B} \subseteq [1:n],\ |\mathcal{B}| = d,\ \mathcal{B} \not\ni j,\ i \in \mathcal{B})$ is the collection of data that can be extracted from the other nodes to regenerate node j.

The closure of all achievable $(\bar{\alpha}, \bar{\beta})$ pairs is the *achievable normalized storage-capacity repair-bandwidth tradeoff region* $\mathcal{R}_{n,k,d,\ell}$ for the (n,k,d,ℓ) secure exact-repair regenerating code problem.

In [19], Shah, Rashmi and Kumar proved the following important achievability result for the general (n,k,d,ℓ) secure regenerating code problem:

$$\left(dT_{k,d,\ell,}^{-1}, T_{k,d,\ell}^{-1}\right) \in \mathcal{R}_{n,k,d,\ell} \tag{3.5}$$

where

$$T_{k,d,\ell} := \sum_{i=\ell+1}^{k} (d+1-i). \tag{3.6}$$

Note that when $\ell = 0$ (no repair-secrecy constraint), $\left(dT_{k,d,0}^{-1}, T_{k,d,0}^{-1}\right)$ recovers the MBR point of the (n,k,d) exact-repair regenerating code problem [9]. It has been shown that the SRK point (3.5) is the *only* corner point of the tradeoff region $\mathcal{R}_{n,k,d,\ell}$ for the cases where we have either $d = 2, 3$ [13], or $d = 4$ [14], or $k = 2$ [13], or $\ell = k-1 = d-1$ [13].

3.3 New Results

Consider the (n,k,d,ℓ) secure exact-repair regenerating code problem (with $\ell \geq 1$), and let

$$\ell^*(k,d) := \min\left\{\ell \geq 1 : T_{k,d,\ell} \leq d + \sqrt{d\ell}\right\}. \tag{3.7}$$

Note that $T_{k,d,\ell}$ is monotone non-increasing with respect to ℓ for any given (k,d) pair, so we have

$$T_{k,d,\ell} \leq d + \sqrt{d\ell}, \quad \forall \ell \geq \ell^*(k,d). \tag{3.8}$$

We have the following two *outer* bounds for the tradeoff region $\mathcal{R}_{n,k,d,\ell}$.

Theorem 2. *For the general (n,k,d,ℓ) secure exact-repair regenerating code problem, any achievable normalized storage-capacity repair-bandwidth pair $(\bar{\alpha},\bar{\beta}) \in \mathcal{R}_{n,k,d,\ell}$ must satisfy:*

$$\bar{\beta} \geq T^{-1}_{k,d,\ell}. \tag{3.9}$$

In addition, when $\ell \geq \ell^*(k,d)$, any achievable normalized storage-capacity repair-bandwidth pair $(\bar{\alpha},\bar{\beta}) \in \mathcal{R}_{n,k,d,\ell}$ must also satisfy:

$$\bar{\alpha} \geq dT^{-1}_{k,d,\ell}. \tag{3.10}$$

(Conversely, any $(\bar{\alpha},\bar{\beta})$ satisfying (3.9) and (3.10) is achievable.)

While the proof of (3.9) is straightforward, the proof of (3.10) is long and technical. We shall defer the proof to Section 3.5. Combining (3.9) and (3.10) proves that the SRK point (3.5) is the *only* corner point of the tradeoff region $\mathcal{R}_{n,k,d,\ell}$ when $\ell \geq \ell^*(k,d)$. It is straightforward to verify that the lower bound $\ell^*(k,d) \leq k-1$ for any (k,d) pair and $\ell^*(k,d) = 1$ for $d \in [2:4]$. Therefore, Theorem 2 includes all previous results from [13] and [14] as special cases.

Next, we shift our attention to the cases where $1 \leq \ell < \ell^*(k,d)$. To see how the tradeoff region $\mathcal{R}_{n,k,d,\ell}$ may look like in this case, let us begin with the following

achievability results for the (n, k, d, ℓ) secure exact-repair regenerating code problem with $k = d = n - 1$.

Theorem 3. *For any $t \in [2 : n - \ell]$, we have*

$$(\bar{\alpha}_t, \bar{\beta}_t) \in \mathcal{R}_{n,n-1,n-1,\ell} \tag{3.11}$$

where

$$(t-1)\bar{\alpha}_t = (n-1)\bar{\beta}_t := \begin{pmatrix} n-1 \\ t-1 \end{pmatrix} \Big/ \begin{pmatrix} n-\ell \\ t \end{pmatrix}. \tag{3.12}$$

The proof is based on a new coding scheme, which we shall describe in the next section. Note that when $\ell = 1$, $(\bar{\alpha}_t, \bar{\beta}_t)$ can be simplified as:

$$(\bar{\alpha}_t, \bar{\beta}_t) = \left(\frac{t}{(t-1)(n-t)}, \frac{t}{(n-1)(n-t)} \right). \tag{3.13}$$

In this case, when $t = 2$, $(\bar{\alpha}_t, \bar{\beta}_t)$ coincides with the SRK point (3.5) with $k = d = n - 1$ and $\ell = 1$. Furthermore, note that $\bar{\beta}_t$ is monotone increasing with t, and $\bar{\alpha}_t$ is monotone decreasing with t for any $t \in [2 : n - 1]$ such that $t^2 + t < n$. Thus, *no* pairs of points from the set $\{(\bar{\alpha}_2, \bar{\beta}_2), \ldots, (\bar{\alpha}_{t+1}, \bar{\beta}_{t+1})\}$ dominate each other for any $t \in [2 : n - 1]$ such that $t^2 + t < n$. For example, when $n = 7$, a second achievability point $(\bar{\alpha}_3, \bar{\beta}_3) = \left(\frac{3}{8}, \frac{1}{8}\right)$ emerges in addition to the SKR point $(\bar{\alpha}_2, \bar{\beta}_2) = \left(\frac{2}{5}, \frac{1}{15}\right)$. When $n = 13$, a third achievability point $(\bar{\alpha}_4, \bar{\beta}_4) = \left(\frac{4}{27}, \frac{1}{27}\right)$ emerges in addition to the points $(\bar{\alpha}_3, \bar{\beta}_3) = \left(\frac{3}{20}, \frac{1}{40}\right)$ and $(\bar{\alpha}_2, \bar{\beta}_2) = \left(\frac{2}{11}, \frac{1}{66}\right)$. Therefore, for the $(n, n-1, n-1, 1)$ secure exact-repair regenerating code problem, the SRK point *cannot* be the only corner point when $n \geq 7$.

Next, we show that both $(\bar{\alpha}_2, \bar{\beta}_2)$ and $(\bar{\alpha}_3, \bar{\beta}_3)$ are *optimal* tradeoff points for the $(n, n-$

$1, n-1, 1)$ secure exact-repair regenerating code problem when $n \geq 7$, so in this case the tradeoff region must have *multiple* corner points.

Theorem 4. *For the* $(n, n-1, n-1, 1)$ *secure exact-repair regenerating code problem with* $n \geq 7$*, any achievable normalized storage-capacity repair-bandwidth pair* $(\bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,n-1,n-1,1}$ *must satisfy:*

$$n\bar{\alpha} + \frac{(n-1)(n-6)}{2}\bar{\beta} \geq 3. \tag{3.14}$$

Note that both

$$(\bar{\alpha}_2, \bar{\beta}_2) = \left(\frac{2}{n-2}, \frac{2}{(n-1)(n-2)}\right)$$
$$\text{and} \quad (\bar{\alpha}_3, \bar{\beta}_3) = \left(\frac{3}{2(n-3)}, \frac{3}{(n-1)(n-3)}\right)$$

satisfy the inequality (3.14) with *equalities* and hence *cannot* be dominated by a single achievable tradeoff point.

Finally, we focus on the $(n, n-1, n-1, 1)$ problem with $n = 7$ and show that the tradeoff region has exactly *two* corner points at $(\bar{\alpha}_2, \bar{\beta}_2)$ and $(\bar{\alpha}_3, \bar{\beta}_3)$.

Theorem 5. *For the* $(7, 6, 6, 1)$ *secure exact-repair regenerating code problem, any achievable normalized storage-capacity repair-bandwidth pair* $(\bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{7,6,6,1}$ *must satisfy:*

$$\bar{\alpha} \geq \frac{3}{8}. \tag{3.15}$$

Therefore, the tradeoff region $\mathcal{R}_{7,6,6,1}$ *is given by:*

$$\mathcal{R}_{7,6,6,1} = \left\{(\bar{\alpha}, \bar{\beta}) : \bar{\beta} \geq \frac{1}{15}, 7\bar{\alpha} + 3\bar{\beta} \geq 3, \bar{\alpha} \geq \frac{3}{8}\right\}. \tag{3.16}$$

The proof of (3.14) and (3.15) can be found in Section 3.5.

3.4 A New $(n, n-1, n-1, \ell)$ Code Construction

In this section, we provide a code construction based on the layered exact-repair regenerating codes proposed in [20], which leads to the achievability of the tradeoff points given in Theorem 3.

Fix a parameter t, and consider the following scheme. There are a total of $B = \binom{n-\ell}{t}(t-1)$ information symbols, denoted as M, and there are a total of $R = \binom{n}{t}(t-1) - B$ random symbols, denoted as K. Assume that an eavesdropper has access to the repair messages to an arbitrary set of ℓ nodes, the collection of which is denoted as E.

We first encode the $(R+B) = \binom{n}{t}(t-1)$ symbols (in a finite field $\mathbb{F}_q$ such that $q \geq 2(R+B)$, and q will be used as the basis of the entropy function), into the parity check symbols of an $(2(R+B), R+B)$ systematic MDS code. These $R+B$ symbols are broken into a total of $\binom{n}{t}$ parity groups, each with $(t-1)$ symbols, and each parity group is associated with a subset $\mathcal{A}$ of $[1:n]$ of cardinality t.

Next, we expand each parity group by introducing one additional parity symbol which can be the simple linear sum of them, in the same finite field as earlier (or even in the binary field, assuming the original one is an extension of the binary field). These t symbols are then distributed into the subset of nodes associated with this parity group, one symbol to each node.

We need to show that the following three conditions are satisfied:

1) Reconstruction with any $n-1$ nodes. This is trivial since in each parity group, at most one of them is in the failed node, and thus the contents of the parity group can be recovered. This also implies that

$$\alpha_t = \binom{n-1}{t-1}. \tag{3.17}$$

2) Repair with the remaining $n-1$ nodes. Assume without loss of generality that node 1 fails. Then, to repair the symbol in the parity group associated with each $\mathcal{A}$ such that $1 \in \mathcal{A}$ and $|\mathcal{A}| = t$, we can send from the remaining nodes all the other symbols in this parity group. The total transmission is thus given by:

$$(n-1)\beta_t = (t-1)\alpha_t. \tag{3.18}$$

3) Security against any eavesdropper on ℓ nodes. We need to show that

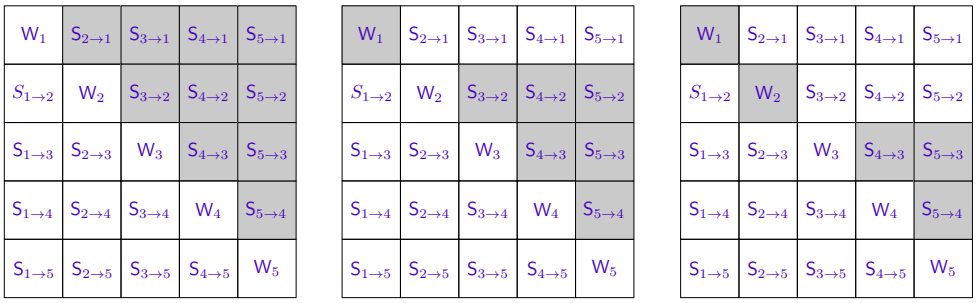
$$I(\mathsf{M};\mathsf{E}) = 0. \tag{3.19}$$

This follows from

$$\begin{aligned} I(\mathsf{M};\mathsf{E}) &= H(\mathsf{E}) - H(\mathsf{E}|\mathsf{M}) \\ &= H(\mathsf{E}) - H(\mathsf{E}|\mathsf{M}) + H(\mathsf{E}|\mathsf{M},\mathsf{K}) \\ &= H(\mathsf{E}) - I(\mathsf{E};\mathsf{K}|\mathsf{M}) \\ &= H(\mathsf{E}) - H(\mathsf{K}|\mathsf{M}) + H(\mathsf{K}|\mathsf{M},\mathsf{E}) \\ &= H(\mathsf{E}) - R + H(\mathsf{K}|\mathsf{M},\mathsf{E}). \end{aligned}$$

All the parity groups that have symbols in the compromised nodes are completely revealed by accessing E, and conversely all the symbols in E can be generated by these parity groups alone. A total of $\binom{n}{t} - \binom{n-\ell}{t}$ parity groups are exposed, implying that

$$H(\mathsf{E}) \leq R.$$



(a) $\mathsf{L}_{0,4}$ (b) $\mathsf{L}_{1,4}$ (c) $\mathsf{L}_{2,4}$

Figure 3.5: Illustration of $\mathsf{L}_{0,4}$, $\mathsf{L}_{1,4}$ and $\mathsf{L}_{2,4}$ in the repair diagram for $n = 5$.

It only remains to show that

$$H(\mathsf{K}|\mathsf{M}, \mathsf{E}) = 0. \tag{3.20}$$

which follows from the fact that given the eavesdropper's information and the message M, the random symbols K can be completely recovered. This can be seen as follows: there are a total of R symbols (after removing the simple sums in each parity group) from E that were original parity symbols of the $(2(R+B), R+B)$ MDS code, but any $R+B$ codeword symbols can be used to recover the original information (M, R), which we indeed have together with the B information symbols.

Normalizing α_t and β_t by B proves the achievability of the tradeoff points given in Theorem 3.

3.5 Proof of the Converse Results

3.5.1 Proof of Theorem 2

Let us first outline the main ingredients for proving the inequalities (3.9) and (3.10).

1) *Total number of nodes.* To prove the inequalities (3.9) and (3.10), let us first note that these two inequalities are *independent* of the total number of storage nodes n in

the system. In our proof, we only need to consider the cases where $n = d + 1$. For the cases where $n > d + 1$, since any subsystem consisting of $d + 1$ out of the total n storage nodes must give rise to a $(d + 1, k, d, \ell)$ secure exact-repair regenerating code problem. Therefore, these two inequalities as *outer* bounds must apply as well. When $n = d + 1$, any repair group $\mathcal{B}$ of size d is uniquely determined by the node j to be repaired, i.e., $\mathcal{B} = [1 : n] \setminus \{j\}$, and hence can be dropped from the notation $\mathsf{S}_{i\to j}^{\mathcal{B}}$ without causing any confusion.

2) *Code symmetry.* Due to the built-in *symmetry* of the problem, to prove the inequalities (3.9) and (3.10), we only need to consider the so-called *symmetrical* codes [8] for which the joint entropy of any subset of random variables from

$$(\mathsf{M}, \mathsf{K}, (\mathsf{W}_i : i \in [1 : n]), (\mathsf{S}_{i\to j} : i, j \in [1 : n], i \neq j))$$

remains *unchanged* under any permutation over the storage-node indices.

3) *Key collections of random variables.* Focusing on the symmetrical $(n = d + 1, d, N, K, T, S)$ codes, the following collections of random variables play a key

role in our proof:

$$\mathsf{W}_{\mathcal{A}} := (\mathsf{W}_i : i \in \mathcal{A}), \quad \mathcal{A} \subseteq [1:n] \tag{3.21}$$

$$\mathsf{S}_{i\to\mathcal{B}} := (\mathsf{S}_{i\to j} : j \in \mathcal{B}),$$
$$i \in [1:n],\ \mathcal{B} \subseteq [1:n] \setminus \{i\} \tag{3.22}$$

$$\mathsf{S}_{\mathcal{B}\to j} := (\mathsf{S}_{i\to j} : i \in \mathcal{B}),$$
$$j \in [1:n],\ \mathcal{B} \subseteq [1:n] \setminus \{j\} \tag{3.23}$$

$$\mathsf{S}_{\to j} := \mathsf{S}_{[1:j-1]\cup[j+1:n]\to j}, \quad j \in [1:n] \tag{3.24}$$

$$\mathsf{S}_{\to\mathcal{B}} := (\mathsf{S}_{\to j} : j \in \mathcal{B}), \quad \mathcal{B} \subseteq [1:n] \tag{3.25}$$

$$\underline{\mathsf{S}}_{\to j} := \mathsf{S}_{[1:j-1]\to j}, \quad j \in [1:n] \tag{3.26}$$

$$\underline{\mathsf{S}}_{\to\mathcal{B}} := (\underline{\mathsf{S}}_{\to j} : j \in \mathcal{B}), \quad \mathcal{B} \subseteq [1:n] \tag{3.27}$$

$$\overline{\mathsf{S}}_{\to j} := \mathsf{S}_{[j+1:n]\to j}, \quad j \in [1:n] \tag{3.28}$$

$$\overline{\mathsf{S}}_{\to\mathcal{B}} := (\overline{\mathsf{S}}_{\to j} : j \in \mathcal{B}), \quad \mathcal{B} \subseteq [1:n] \tag{3.29}$$

$$\mathsf{L}_{t,s} := (\mathsf{W}_{[1:t]}, \overline{\mathsf{S}}_{\to[t+1:s]}),$$
$$s \in [1:n],\ t \in [0:s]. \tag{3.30}$$

In particular, the collection $\mathsf{L}_{t,s}$ defined in (3.30) was first identified in [12] for proving that separate encoding can achieve the MBR point for multilevel diversity coding with regeneration. As we shall see, here it also plays a key role in our proof of (3.9) and (3.10). Fig. 3.5 illustrates the structure of $\mathsf{L}_{0,4}$, $\mathsf{L}_{1,4}$ and $\mathsf{L}_{2,4}$ in the *repair diagram* introduced by Duursma [11] for $n = 5$.

An important part of the proof is to understand the relations between the collections of random variables defined above, and to use them to derive the desired converse results. We have the following key lemmas, whose proof can be found in the Appendix.

Lemma 1. *For any* $(n = d+1, k, d, N, K, T, S)$ *code that satisfies the node-regeneration requirement* (3.3)*,* $(\underline{\mathsf{S}}_{\to[t+1:s]}, \mathsf{W}_{[t+1:s]})$ *is a function of* $\mathsf{L}_{t,s}$ *for any* $s \in [1:n]$ *and* $t \in [0: s-1]$.

Lemma 2. *For any symmetrical* $(n = d+1, k, d, N, K, T, S)$ *code, we have*

$$\begin{aligned} &H(\mathsf{S}_{1\to[2:p+1]}) + H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+1]\to r+1}) \\ &\quad \geq H(\mathsf{S}_{1\to[2:p]}) + H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+2]\to r+1}) \end{aligned} \tag{3.31}$$

for any $t \in [1:2]$*,* $r \in [2:k-1]$*,* $p \in [1: r-t+1]$*, and* $q \in [0: d-r-1]$*. It follows that*

$$H(\mathsf{L}_{t,j}) + T_{k,d,j} m^{-1} H(\mathsf{S}_{1\to[2:m+1]}) \geq H(\mathsf{L}_{t,k}) \tag{3.32}$$

for any $t \in [1:2]$*,* $j \in [2:k]$*, and* $m \in [1: j-t+1]$.

Lemma 3. *For any symmetrical* $(n = d+1, k, d, N, K, T, S)$ *code that satisfies the node-regeneration requirement* (3.3)*, we have*

$$\begin{aligned} &\frac{d-t}{n-j} H(\mathsf{L}_{1,j}, \mathsf{S}_{j\to 1}) + H(\mathsf{L}_{1,t}) \\ &\quad \geq \frac{d-t}{n-j} H(\mathsf{L}_{1,j-1}, \mathsf{S}_{j\to 1}) + H(\mathsf{L}_{1,t+1}) \end{aligned} \tag{3.33}$$

for any $j \in [2:k-1]$ *and* $t \in [j: k-1]$*. It follows that*

$$\begin{aligned} &H(\mathsf{L}_{1,j}, \mathsf{S}_{j\to 1}) + (n-j) T_{k,d,m}^{-1} H(\mathsf{L}_{1,m}) \\ &\quad \geq H(\mathsf{L}_{1,j-1}, \mathsf{S}_{j\to 1}) + (n-j) T_{k,d,m}^{-1} H(\mathsf{L}_{1,k}) \end{aligned} \tag{3.34}$$

for any $j \in [2:k-1]$ *and* $m \in [j: k-1]$.

The inequality (3.9) can now be proved as follows:

$$
\begin{aligned}
B &= H(\mathsf{M}) \\
&\overset{(a)}{=} H(\mathsf{M}|\mathsf{S}_{\to[1:\ell]}) \\
&\overset{(b)}{=} H(\mathsf{M}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}) \\
&\leq H(\mathsf{M}, \overline{\mathsf{S}}_{\to[\ell+1:k]}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}) \\
&= H(\mathsf{M}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}, \overline{\mathsf{S}}_{\to[\ell+1:k]}) \\
&\quad + H(\bar{\mathsf{S}}_{\to[\ell+1:k]}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}) \\
&= H(\mathsf{M}|\mathsf{S}_{\to[1:\ell]}, \mathsf{L}_{\ell,k}) + H(\bar{\mathsf{S}}_{\to[\ell+1:k]}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}) \\
&\overset{(c)}{=} H(\mathsf{M}|\mathsf{S}_{\to[1:\ell]}, \mathsf{L}_{\ell,k}, \mathsf{W}_{[\ell+1:k]}) \\
&\quad + H(\bar{\mathsf{S}}_{\to[\ell+1:k]}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}) \\
&= H(\mathsf{M}|\mathsf{S}_{\to[1:\ell]}, \overline{\mathsf{S}}_{\to[\ell+1:k]}, \mathsf{W}_{[1:k]}) \\
&\quad + H(\bar{\mathsf{S}}_{\to[\ell+1:k]}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}) \\
&\overset{(d)}{=} H(\bar{\mathsf{S}}_{\to[\ell+1:k]}|\mathsf{S}_{\to[1:\ell]}, \mathsf{W}_{[1:\ell]}) \\
&\leq H(\bar{\mathsf{S}}_{\to[\ell+1:k]}) \\
&\overset{(d)}{\leq} T_{k,d,l}\beta
\end{aligned}
$$

where (a) follows from the repair-secrecy constraint (3.4); (b) follows from the fact that $\mathsf{W}_{[1:\ell]}$ is a function of $\mathsf{S}_{\to[1:\ell]}$ due to the node-regeneration constraint (3.3); (c) follows from the fact that $\mathsf{W}_{[\ell+1:k]}$) is a function of $\mathsf{L}_{\ell,k}$ by Lemma 1; (d) follows from the fact that $H(\mathsf{M}|\mathsf{S}_{\to[1:\ell]}, \overline{\mathsf{S}}_{\to[\ell+1:k]}, \mathsf{W}_{[1:k]}) = 0$ due to the message-recovery constraint (3.2); and (e) follows from the bandwidth constraint on the repair messages. Normalizing both sides by B completes the proof of (3.9).

To prove the inequality (3.10), we shall consider the cases where $T_{k,d,\ell} \leq d$ and $d \leq T_{k,d,\ell} \leq d + \sqrt{d\ell}$ separately.

Case 1: $T_{k,d,\ell} \leq d$. In this case, we have

$$
\begin{aligned}
& T_{k,d,\ell}\alpha + dH(\mathsf{S}_{\rightarrow[1:\ell]}) \\
&= T_{k,d,\ell}\left(\alpha + H(\mathsf{S}_{\rightarrow[1:\ell]})\right) + (d - T_{k,d,\ell})\, H(\mathsf{S}_{\rightarrow[1:\ell]}) \\
&\overset{(a)}{=} T_{k,d,\ell}\left(\alpha + H(\mathsf{S}_{\rightarrow[2:\ell+1]}\right) + (d - T_{k,d,\ell})\, H(\mathsf{S}_{\rightarrow[1:\ell]}) \\
&\overset{(b)}{\geq} T_{k,d,\ell}\left(H(\mathsf{W}_1) + H(\mathsf{S}_{\rightarrow[2:\ell+1]})\right) + (d - T_{k,d,\ell})\, H(\mathsf{S}_{\rightarrow[1:\ell]}) \\
&\overset{(c)}{=} T_{k,d,\ell}\left(H(\mathsf{W}_1, \mathsf{S}_{1\rightarrow[2:\ell+1]}) + H(\mathsf{S}_{\rightarrow[2:\ell+1]})\right) \\
&\quad + (d - T_{k,d,\ell})\, H(\mathsf{S}_{\rightarrow[1:\ell]}) \\
&\overset{(d)}{\geq} T_{k,d,\ell}\left(H(\mathsf{W}_1, \mathsf{S}_{\rightarrow[2:\ell+1]}) + H(\mathsf{S}_{1\rightarrow[2:\ell+1]})\right) \\
&\quad + (d - T_{k,d,\ell})\, H(\mathsf{S}_{\rightarrow[1:\ell]}) \\
&\overset{(e)}{=} T_{k,d,\ell}\left(H(\mathsf{W}_1, \mathsf{S}_{\rightarrow[2:\ell+1]}) + T_{k,d,\ell+1}\ell^{-1}H(\mathsf{S}_{1\rightarrow[2:\ell+1]})\right) \\
&\quad + (d - T_{k,d,\ell})\left(H(\mathsf{S}_{\rightarrow[1:\ell]}) + T_{k,d,\ell}\ell^{-1}H(\mathsf{S}_{1\rightarrow[2:\ell+1]})\right) \\
&\overset{(f)}{\geq} T_{k,d,\ell}\left(H(\mathsf{L}_{1,\ell+1}) + T_{k,d,\ell+1}\ell^{-1}H(\mathsf{S}_{1\rightarrow[2:\ell+1]})\right) \\
&\quad + (d - T_{k,d,\ell})\left(H(\mathsf{L}_{1,\ell}) + T_{k,d,\ell}\ell^{-1}H(\mathsf{S}_{1\rightarrow[2:\ell+1]})\right)
\end{aligned} \tag{3.35}
$$

where (a) follows from the fact that

$$
H(\mathsf{S}_{\rightarrow[1:\ell]}) = H(\mathsf{S}_{\rightarrow[2:\ell+1]}) \tag{3.36}
$$

due to the symmetrical code that we consider; (b) is due to the storage-capacity constraint $H(\mathsf{W}_1) \leq \alpha$; (c) is due to the fact that $\mathsf{S}_{1\rightarrow[2:\ell+1]}$ is a function of W_1; (d) follows from the

fact that

$$H(\mathsf{W}_1,\mathsf{S}_{1\to[2:\ell+1]}) + H(\mathsf{S}_{\to[2:\ell+1]}) \\ \geq H(\mathsf{W}_1, \mathsf{S}_{\to[2:\ell+1]}) + H(\mathsf{S}_{1\to[2:\ell+1]}) \tag{3.37}$$

due to the submodularity of the entropy function; (e) follows from the fact that

$$T_{k,d,\ell+1}\ell^{-1} + (d - T_{k,d,\ell})\,\ell^{-1} = 1;$$

and (f) follows from the facts that

$$H(\mathsf{W}_1, \mathsf{S}_{\to[2:\ell+1]}) \geq H(\mathsf{W}_1, \overline{\mathsf{S}}_{\to[2:\ell+1]}) = H(\mathsf{L}_{1,\ell+1}) \tag{3.38}$$

$$H(\mathsf{S}_{\to[1:\ell]}) = H(\mathsf{W}_1, \mathsf{S}_{\to[1:\ell]}) \\ \geq H(\mathsf{W}_1, \overline{\mathsf{S}}_{\to[2:\ell]}) = H(\mathsf{L}_{1,\ell}). \tag{3.39}$$

Applying (3.32) with $(t, j, m) = (1, \ell + 1, \ell)$ and $(t, j, m) = (1, \ell, \ell)$, respectively gives:

$$H(\mathsf{L}_{1,\ell+1}) + T_{k,d,\ell+1}\ell^{-1}H(\mathsf{S}_{1\to[2:\ell+1]}) \geq H(\mathsf{L}_{1,k}) \tag{3.40}$$

$$H(\mathsf{L}_{1,\ell}) + T_{k,d,\ell+1}\ell^{-1}H(\mathsf{S}_{1\to[2:\ell+1]}) \geq H(\mathsf{L}_{1,k}). \tag{3.41}$$

Substituting (3.40) and (3.41) into (3.35) gives:

$$\begin{aligned}
&T_{k,d,\ell}\alpha + dH(\mathsf{S}_{\to[1:\ell]}) \geq dH(\mathsf{L}_{1,k}) \\
&\overset{(a)}{=} dH(\mathsf{L}_{1,k}, \mathsf{W}_{[2:k]}, \underline{\mathsf{S}}_{\to[2:k]}) = dH(\mathsf{W}_{[1:k]}, \mathsf{S}_{\to[2:k]}) \\
&\overset{(b)}{=} dH(\mathsf{W}_{[1:k]}, \mathsf{M}, \mathsf{S}_{\to[2:k]}) \geq dH(\mathsf{M}, \mathsf{S}_{\to[2:\ell+1]}) \\
&= dH(\mathsf{S}_{\to[2:\ell+1]}) + dH(\mathsf{M}|\mathsf{S}_{\to[2:\ell+1]}) \\
&\overset{(c)}{=} dH(\mathsf{S}_{\to[2:\ell+1]}) + dH(\mathsf{M}) \overset{(d)}{=} dH(\mathsf{S}_{\to[1:\ell]}) + dB
\end{aligned}$$

where (a) follows from the fact that $(\mathsf{W}_{[2:k]}, \underline{\mathsf{S}}_{\to[2:k]})$ is a function of $\mathsf{L}_{1,k}$ by Lemma 1; (b) follows from the fact that M is a function of $\mathsf{W}_{[1:k]}$ due to the message-recover constraint (3.2); (c) follows from the repair-secrecy constraint (3.4); and (d) follows again from (3.36) due to the symmetrical code that we consider. Canceling $dH(\mathsf{S}_{\to[1:\ell]})$ from both sides of the inequality completes the proof of (3.10) for the cases where $T_{k,d,\ell} \leq d$.

Case 2: $d \leq T_{k,d,\ell} \leq d + \sqrt{d\ell}$. Note that if $k = \ell + 1$, we have $T_{k,d,\ell} = d - \ell < d$. Therefore, in this case we must have $k \geq \ell + 2 \geq 3$. In addition, let

$$q := 1 + (d - \ell)T_{k,d,\ell+1}^{-1}$$

and we have

$$\begin{aligned}
&d - (T_{k,d,\ell} - d)\, q \\
&\quad = T_{k,d,\ell+1}^{-1} \left(-T_{k,d,\ell}^2 + 2dT_{k,d,\ell} - d(d-\ell)\right) \geq 0.
\end{aligned}$$

It follows that

$$
\begin{aligned}
&T_{k,d,\ell}\alpha + dH(\mathsf{S}_{\to[1:\ell]}) \\
&= d\left(\alpha + H(\mathsf{S}_{\to[1:\ell]})\right) + (T_{k,d,\ell} - d)\,\alpha \\
&\overset{(a)}{=} d\left(\alpha + H(\mathsf{S}_{\to[2:\ell+1]})\right) + (T_{k,d,\ell} - d)\,\alpha \\
&\overset{(b)}{\geq} d\left(H(\mathsf{W}_1) + H(\mathsf{S}_{\to[2:\ell+1]})\right) + (T_{k,d,\ell} - d)\,\alpha \\
&\overset{(c)}{=} d\left(H(\mathsf{W}_1, \mathsf{S}_{1\to[2:\ell+1]}) + H(\mathsf{S}_{\to[2:\ell+1]})\right) + (T_{k,d,\ell} - d)\,\alpha \\
&\overset{(d)}{\geq} d\left(H(\mathsf{W}_1, \mathsf{S}_{\to[2:\ell+1]}) + H(\mathsf{S}_{1\to[2:\ell+1]})\right) + (T_{k,d,\ell} - d)\,\alpha \\
&\overset{(e)}{\geq} d\left(H(\mathsf{L}_{1,\ell+1}) + H(\mathsf{S}_{1\to[2:\ell+1]})\right) + (T_{k,d,\ell} - d)\,\alpha \\
&= (d - (T_{k,d,\ell} - d)\,q)\,H(\mathsf{L}_{1,\ell+1}) \\
&\quad + (T_{k,d,\ell} - d)\,(qH(\mathsf{L}_{1,\ell+1}) + \alpha) + dH(\mathsf{S}_{1\to[2:\ell+1]})
\end{aligned} \tag{3.42}
$$

where (a) follows from (3.36) due to the symmetrical code that we consider; (b) is due to the storage-capacity constraint $H(\mathsf{W}_1) \leq \alpha$; (c) is due to the fact that $\mathsf{S}_{1\to[2:\ell+1]}$ is a function of W_1; (d) follows from (3.37) due to the submodularity of the entropy function; and (e) follows from (3.38).

The first term on the right-hand side of (3.42) can be further bounded from below by the fact that $\mathsf{L}_{2,\ell+1}$ is a function of $\mathsf{L}_{1,\ell+1}$ by Lemma 1, so we have

$$
H(\mathsf{L}_{1,\ell+1}) \geq H(\mathsf{L}_{2,\ell+1}). \tag{3.43}
$$

To bound from below the second term on the right-hand side of (3.42), note that

$$
\begin{aligned}
& qH(\mathsf{L}_{1,\ell+1}) + \alpha \\
& = H(\mathsf{L}_{1,\ell+1}) + (d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{1,\ell+1}) + \alpha \\
& \overset{(a)}{=} H(\mathsf{L}_{1,\ell+1}, \mathsf{S}_{\ell+1\to 1}) + (d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{1,\ell+1}) + \alpha \\
& \overset{(b)}{\geq} H(\mathsf{L}_{1,\ell}, \mathsf{S}_{\ell+1\to 1}) + (d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{1,k}) + \alpha \\
& \overset{(c)}{\geq} H(\mathsf{L}_{1,\ell}, \mathsf{S}_{\ell+1\to 1}) + (d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{1,k}) + H(\mathsf{W}_{\ell+1}) \\
& \overset{(d)}{=} H(\mathsf{L}_{1,\ell}, \mathsf{S}_{\ell+1\to 1}) + H(\mathsf{W}_{\ell+1}, \mathsf{S}_{\ell+1\to[1:\ell]}) + \\
& \quad (d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{1,k}) \\
& \overset{(e)}{\geq} H(\mathsf{L}_{1,\ell}, \mathsf{W}_{\ell+1}, \mathsf{S}_{\ell+1\to 1}) + H(\mathsf{S}_{\ell+1\to[1:\ell]}) + \\
& \quad (d-\ell)T_{d,\ell+1,k}^{-1} H(\mathsf{L}_{1,k}) \\
& \overset{(f)}{=} H(\mathsf{L}_{1,\ell}, \mathsf{W}_{\ell+1}) + H(\mathsf{S}_{\ell+1\to[1:\ell]}) + (d-l)T_{k,d,l+1}^{-1} H(\mathsf{L}_{1,k}) \\
& \overset{(g)}{=} H(\mathsf{L}_{2,\ell+1}) + H(\mathsf{S}_{1\to[2:\ell+1]}) + (d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{1,k}) \\
& \overset{(h)}{\geq} H(\mathsf{L}_{2,\ell+1}) + H(\mathsf{S}_{1\to[2:\ell+1]}) + (d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{2,k})
\end{aligned} \tag{3.44}
$$

where (a) follows from the fact that $\mathsf{S}_{\ell+1\to 1}$ is a function of $\mathsf{W}_{\ell+1}$, which is in turn a function of $\mathsf{L}_{1,\ell+1}$ by Lemma 1; (b) follows from (3.34) with $(j,m) = (\ell+1, \ell+1)$; (c) is due to the storage-capacity constraint $H(\mathsf{W}_{\ell+1}) \leq \alpha$; (d) follows from the fact that $\mathsf{S}_{\ell+1\to 1}$ is a function of $\mathsf{W}_{\ell+1}$; (e) follows from the fact that

$$
\begin{aligned}
& H(\mathsf{L}_{1,\ell}, \mathsf{S}_{\ell+1\to 1}) + H(\mathsf{W}_{\ell+1}, \mathsf{S}_{\ell+1\to[1:\ell]}) \\
& \qquad \geq H(\mathsf{L}_{1,\ell}, \mathsf{W}_{\ell+1}, \mathsf{S}_{\ell+1\to 1}) + H(\mathsf{S}_{\ell+1\to[1:\ell]})
\end{aligned} \tag{3.45}
$$

due to the submodularity of the entropy function; (f) follows yet again from the fact that $\mathsf{S}_{\ell+1\to 1}$ is a function of $\mathsf{W}_{\ell+1}$; (g) follows from the facts that

$$H(\mathsf{L}_{1,\ell}, \mathsf{W}_{\ell+1}) = H(\mathsf{L}_{2,\ell+1}) \tag{3.46}$$

$$H(\mathsf{S}_{\ell+1\to[1:\ell]}) = H(\mathsf{S}_{1\to[2:\ell+1]}) \tag{3.47}$$

due to the symmetrical code that we consider; and (h) follows from the fact that $\mathsf{L}_{2,k}$ is a function of $\mathsf{L}_{1,k}$ by Lemma 1, so we have

$$H(\mathsf{L}_{1,k}) \geq H(\mathsf{L}_{2,k}). \tag{3.48}$$

Substituting (3.43) and (3.44) into (3.42) gives:

$$\begin{aligned}
&T_{k,d,\ell}\alpha + dH(\mathsf{S}_{\to[1:\ell]}) \\
&\geq \left(d - (T_{k,d,\ell} - d)(d-\ell)T_{k,d,\ell+1}^{-1}\right) H(\mathsf{L}_{2,\ell+1}) + \\
&\quad T_{k,d,\ell} H(\mathsf{S}_{1\to[2:\ell+1]}) + (T_{k,d,\ell} - d)\,(d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{2,k}) \\
&\overset{(a)}{=} T_{k,d,\ell}T_{k,d,\ell+1}^{-1}\ell\left(H(\mathsf{L}_{2,\ell+1}) + T_{k,d,\ell+1}\ell^{-1}H(\mathsf{S}_{1\to[2:\ell+1]}\right) + \\
&\quad (T_{k,d,\ell} - d)\,(d-\ell)T_{k,d,\ell+1}^{-1} H(\mathsf{L}_{2,k}) \\
&\overset{(b)}{\geq} \left(T_{k,d,\ell}T_{k,d,\ell+1}^{-1}\ell + (T_{k,d,\ell} - d)(d-\ell)T_{k,d,\ell+1}^{-1}\right) H(\mathsf{L}_{2,k}) \\
&\overset{(c)}{=} dH(\mathsf{L}_{2,k}) \overset{(d)}{=} dH(\mathsf{L}_{2,k}, \mathsf{W}_{[3:k]}, \underline{\mathsf{S}}_{\to[3:k]}) \\
&= dH(\mathsf{W}_{[1:k]}, \mathsf{S}_{\to[3:k]}) \overset{(e)}{=} dH(\mathsf{W}_{[1:k]}, \mathsf{M}, \mathsf{S}_{\to[3:k]}) \\
&\geq dH(\mathsf{M}, \mathsf{S}_{\to[3:\ell+2]}) = dH(\mathsf{S}_{\to[3:\ell+2]}) + dH(\mathsf{M}|\mathsf{S}_{\to[3:\ell+2]}) \\
&\overset{(f)}{=} dH(\mathsf{S}_{\to[3:\ell+2]}) + dH(\mathsf{M}) \overset{(g)}{=} dH(\mathsf{S}_{\to[1:\ell]}) + dB
\end{aligned}$$

where (a) and (c) follow from the fact that

$$d-(T_{k,d,\ell}-d)(d-\ell)T_{k,d,\ell+1}^{-1}=T_{k,d,\ell}T_{k,d,\ell+1}^{-1}\ell;$$

(b) follows from (3.32) with $(t,j,m)=(2,\ell+1,l)$; (d) follows from the fact that $(\mathsf{W}_{[3:k]},\underline{\mathsf{S}}_{\rightarrow[3:k]})$ is a function of $\mathsf{L}_{2,k}$ by Lemma 1; (e) follows from the fact that M is a function of $\mathsf{W}_{[1:k]}$ due to the message-recover constraint (3.2); (f) is due to the secrecy constraint (3.4); and (g) follows from the fact that

$$H(\mathsf{S}_{\rightarrow[3:\ell+2]})=H(\mathsf{S}_{\rightarrow[1:\ell]}) \tag{3.49}$$

due to the symmetrical code that we consider. Canceling $dH(\mathsf{S}_{\rightarrow[1:\ell]})$ from both sides of the inequality completes the proof of (3.10) for $d\leq T_{k,d,\ell}\leq d+\sqrt{d\ell}$.

3.5.2 Proof of Theorem 4

Assume that $k=d=n-1$ and $\ell=1$. As before, we shall also assume without loss of generality that the codes that we consider here are symmetrical ones.

Let us first show that for any $i\in[1:n-1]$, we have

$$H(\mathsf{S}_{\rightarrow 1})\geq H(\overline{\mathsf{S}}_{\rightarrow i}|\mathsf{W}_{[1:i-1]})+H(\underline{\mathsf{S}}_{\rightarrow i}) \tag{3.50}$$

which can be seen as follows:

$$
\begin{aligned}
& H(\mathsf{W}_{[1:i-1]}) + H(\mathsf{S}_{\to 1}) \\
& \quad \overset{(a)}{=} H(\mathsf{W}_{[1:i-1]}) + H(\mathsf{S}_{\to i}) \\
& \quad \overset{(b)}{=} H(\mathsf{W}_{[1:i-1]}, \underline{\mathsf{S}}_{\to i}) + H(\underline{\mathsf{S}}_{\to i}, \overline{\mathsf{S}}_{\to i}) \\
& \quad \overset{(c)}{\geq} H(\mathsf{W}_{[1:i-1]}, \underline{\mathsf{S}}_{\to i}, \overline{\mathsf{S}}_{\to i}) + H(\underline{\mathsf{S}}_{\to i}) \\
& \quad \geq H(\mathsf{W}_{[1:i-1]}, \overline{\mathsf{S}}_{\to i}) + H(\underline{\mathsf{S}}_{\to i}) \\
& \quad \geq H(\mathsf{W}_{[1:i-1]}) + H(\overline{\mathsf{S}}_{\to i} | \mathsf{W}_{[1:i-1]}) + H(\underline{\mathsf{S}}_{\to i})
\end{aligned}
$$

where (a) follows from the fact that $H(\mathsf{S}_{\to 1}) = H(\mathsf{S}_{\to i})$ due to the symmetrical code that we consider; (b) is due to the fact that $\underline{\mathsf{S}}_{\to i}$ is a function of $\mathsf{W}_{[1:i-1]}$; and (c) is due to the submodularity of the entropy function. Canceling $H(\mathsf{W}_{[1:i-1]})$ from both sides completes the proof of (3.50).

Setting $i = 2, 3$ in (3.50) and by the symmetrical code that we consider, we have

$$
\begin{aligned}
H(\mathsf{S}_{\to 1}) &\geq H(\overline{\mathsf{S}}_{\to 2} | \mathsf{W}_1) + H(\underline{\mathsf{S}}_{\to 2}) \\
&= H(\overline{\mathsf{S}}_{\to 2} | \mathsf{W}_1) + H(\mathsf{S}_{1 \to 2}) \\
&= H(\overline{\mathsf{S}}_{\to 2} | \mathsf{W}_1) + H(\mathsf{S}_{n \to n-1}) \\
&= H(\overline{\mathsf{S}}_{\to 2} | \mathsf{W}_1) + H(\overline{\mathsf{S}}_{\to n-1}) \\
&\geq H(\overline{\mathsf{S}}_{\to 2} | \mathsf{W}_1) + H(\overline{\mathsf{S}}_{\to n-1} | \mathsf{W}_{[1:n-2]}) \qquad (3.51)
\end{aligned}
$$

and

$$\begin{aligned} H(\mathsf{S}_{\to 1}) &\geq H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\underline{\mathsf{S}}_{\to 3}) \\ &= H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\mathsf{S}_{[1:2]\to 3}) \\ &= H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\mathsf{S}_{[n-1:n]\to n-2}) \\ &= H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\overline{\mathsf{S}}_{\to n-2}) \\ &\geq H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\overline{\mathsf{S}}_{\to n-2}|\mathsf{W}_{[1:n-3]}). \end{aligned} \tag{3.52}$$

Adding (3.51) and (3.52) gives:

$$\begin{aligned} &3H(\mathsf{S}_{\to 1}) \\ &\geq H(\mathsf{S}_{\to 1}) + H(\overline{\mathsf{S}}_{\to 2}|\mathsf{W}_1) + H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + \\ &\quad H(\overline{\mathsf{S}}_{\to n-2}|\mathsf{W}_{[1:n-3]}) + H(\overline{\mathsf{S}}_{\to n-1}|\mathsf{W}_{[1:n-2]}). \end{aligned} \tag{3.53}$$

Furthermore, by the repair-bandwidth constraint, we have

$$\begin{aligned} \frac{(n-1)(n-6)}{2}\beta = \sum_{i=3}^{n-4} i\beta &\geq \sum_{i=3}^{n-4} H(\overline{\mathsf{S}}_{\to n-i}) \\ &\geq \sum_{i=3}^{n-4} H(\overline{\mathsf{S}}_{\to n-i}|\mathsf{W}_{[1:n-i-1]}). \end{aligned} \tag{3.54}$$

Adding (3.53) and (3.54) gives:

$$\begin{aligned} &\frac{(n-1)(n-6)}{2}\beta + 3H(\mathsf{S}_{\to 1}) \\ &\geq \sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to n-i}|\mathsf{W}_{[1:n-i-1]}) = \sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) \end{aligned} \tag{3.55}$$

where the last equality follows from the change of variable $i \to n - i$.

To proceed, we shall need the following lemma, whose proof can be found in the Appendix.

Lemma 4. *For any symmetrical* $(n, k = n-1, d = n-1, N, K, T, S)$ *code that satisfies the node-regeneration requirement* (3.3) *and the repair-secrecy constraint* (3.4) *with* $\ell = 1$*, we have*

$$\sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + n\alpha \geq 3H(\mathsf{S}_{\to 1}) + 3B. \tag{3.56}$$

Adding (3.55) and (3.56) gives:

$$\frac{(n-1)(n-6)}{2}\beta + n\alpha + 3H(\mathsf{S}_{\to 1}) \geq 3H(\mathsf{S}_{\to 1}) + 3B.$$

Canceling $3H(\mathsf{S}_{\to 1})$ from both sides and normalizing the remaining terms by B complete the proof of (3.14).

3.5.3 Proof of Theorem 5

Let us first show that

$$3H(\mathsf{S}_{\to 1}) \geq \sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\underline{\mathsf{S}}_{\to 4}) \tag{3.57}$$

which can be seen as follows.

First note that

$$\begin{aligned} H(\mathsf{S}_{\to 1}) &\overset{(a)}{\geq} H(\overline{\mathsf{S}}_{\to 2}|\mathsf{W}_1) + H(\mathsf{S}_{1\to 2}) \\ &\overset{(b)}{=} H(\overline{\mathsf{S}}_{\to 2}|\mathsf{W}_1) + H(\mathsf{S}_{7\to 6}) \\ &= H(\overline{\mathsf{S}}_{\to 2}|\mathsf{W}_1) + H(\overline{\mathsf{S}}_{\to 6}) \\ &\geq H(\overline{\mathsf{S}}_{\to 2}|\mathsf{W}_1) + H(\overline{\mathsf{S}}_{\to 6}|\mathsf{W}_{[1:5]}) \end{aligned} \tag{3.58}$$

where (a) follows from (3.50) with $n = 7$ and $i = 2$; and (b) follows from the fact that $H(\mathsf{S}_{1\to 2}) = H(\mathsf{S}_{7\to 6})$ due to the symmetrical code that we consider. Next, we have

$$\begin{aligned} H(\mathsf{S}_{\to 1}) &\overset{(a)}{\geq} H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\underline{\mathsf{S}}_{\to 3}) \\ &= H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\mathsf{S}_{[1:2]\to 3}) \\ &\overset{(b)}{=} H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\mathsf{S}_{[6:7]\to 5}) \\ &= H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\overline{\mathsf{S}}_{\to 5}) \\ &\geq H(\overline{\mathsf{S}}_{\to 3}|\mathsf{W}_{[1:2]}) + H(\overline{\mathsf{S}}_{\to 5}|\mathsf{W}_{[1:4]}) \end{aligned} \tag{3.59}$$

where (a) follows from (3.50) with $n = 7$ and $i = 3$; and (b) follows from the fact that $H(\mathsf{S}_{[1:2]\to 3}) = H(\mathsf{S}_{[6:7]\to 5})$ due to the symmetrical code that we consider. Finally, setting $n = 7$ and $i = 4$ in (3.50) gives

$$H(\mathsf{S}_{\to 1}) \geq H(\overline{\mathsf{S}}_{\to 4}|\mathsf{W}_{[1:3]}) + H(\underline{\mathsf{S}}_{\to 4}). \tag{3.60}$$

Adding (3.58)–(3.60) completes the proof of (3.57).

To proceed, we shall consider the cases where $\alpha \geq H(\mathsf{S}_{[2:4]\to 1})$ and $\alpha \leq H(\mathsf{S}_{[2:4]\to 1})$, separately.

Case 1: $\alpha \geq H(\mathsf{S}_{[2:4]\to 1})$. In this case, we have

$$
\begin{aligned}
&8\alpha + 3H(\mathsf{S}_{\to 1}) \\
&\geq 7\alpha + 3H(\mathsf{S}_{\to 1}) + H(\mathsf{S}_{[2:4]\to 1}) \\
&\overset{(a)}{\geq} 7\alpha + \sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\underline{\mathsf{S}}_{\to 4}) + H(\mathsf{S}_{[2:4]\to 1}) \\
&= 7\alpha + \sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\mathsf{S}_{[1:3]\to 4}) + H(\mathsf{S}_{[2:4]\to 1}) \\
&\overset{(b)}{=} 7\alpha + \sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\mathsf{S}_{[5:7]\to 1}) + H(\mathsf{S}_{[2:4]\to 1}) \\
&\geq 7\alpha + \sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\mathsf{S}_{\to 1}) \\
&= 7\alpha + \sum_{i=1}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) \\
&\overset{(c)}{\geq} 3H(\mathsf{S}_{\to 1}) + 3B
\end{aligned}
$$

where (a) follows from (3.57); (b) follows from the fact that $H(\mathsf{S}_{[1:3]\to 4}) = H(\mathsf{S}_{[5:7]\to 1})$ due to the symmetrical code that we consider; and (c) follows from Lemma 4 with $n = 7$. Canceling $3H(\mathsf{S}_{\to 1})$ from both sides and normalizing the remaining terms by B complete the proof of (3.15) for the cases where $\alpha \geq H(\mathsf{S}_{[2:4]\to 1})$.

Case 2: $\alpha \leq H(\mathsf{S}_{[2:4]\to 1})$. Note that in this case, by node-capacity constraint and the symmetry of the code that we consider, we have

$$H(\mathsf{W}_1) \leq \alpha \leq H(\mathsf{S}_{[2:4]\to 1}) = H(\mathsf{S}_{[1:3]\to 4}) = H(\underline{\mathsf{S}}_{\to 4}). \tag{3.61}$$

It follows that

$$
\begin{aligned}
&\sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\underline{\mathsf{S}}_{\to 4}) - H(\mathsf{W}_{[1:6]}) \\
&= \sum_{i=2}^{6} \big(H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) - H(\mathsf{W}_i|\mathsf{W}_{[1:i-1]})\big) \\
&\quad + (H(\underline{\mathsf{S}}_{\to 4}) - H(\mathsf{W}_1)) \\
&\overset{(a)}{\geq} \sum_{i=2}^{6} \big(H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) - H(\mathsf{W}_i|\mathsf{W}_{[1:i-1]})\big) \\
&\overset{(b)}{=} \sum_{i=2}^{6} \big(H(\overline{\mathsf{S}}_{\to i}, \mathsf{W}_i|\mathsf{W}_{[1:i-1]}) - H(\mathsf{W}_i|\mathsf{W}_{[1:i-1]})\big) \\
&= \sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i]}) \\
&\geq \sum_{i=2}^{5} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i]}) \\
&= \sum_{i=2}^{5} H(\mathsf{S}_{[i+1:7]\to i}|\mathsf{W}_{[1:i]}) \\
&\geq \sum_{i=2}^{5} H(\mathsf{S}_{[i+1:6]\to i}|\mathsf{W}_{[1:i]}) \\
&= \sum_{i=2}^{5} \sum_{j=i+1}^{6} H(\mathsf{S}_{j\to i}|\mathsf{W}_{[1:i]}, \mathsf{S}_{[i+1:j-1]\to i}) \\
&\overset{(c)}{\geq} \sum_{i=2}^{5} \sum_{j=i+1}^{6} H(\mathsf{S}_{j\to i}|\mathsf{W}_{[1:j-1]}) \\
&= \sum_{j=3}^{6} \sum_{i=2}^{j-1} H(\mathsf{S}_{j\to i}|\mathsf{W}_{[1:j-1]}) \\
&\geq \sum_{j=3}^{6} H(\mathsf{S}_{j\to [2:j-1]}|\mathsf{W}_{[1:j-1]}) \\
&= \sum_{j=3}^{6} H(\mathsf{S}_{j\to [2:j-1]}) - \sum_{j=3}^{6} I(\mathsf{S}_{j\to [2:j-1]}; \mathsf{W}_{[1:j-1]}) \\
&\overset{(d)}{\geq} \sum_{j=3}^{6} H(\mathsf{S}_{j\to [2:j-1]}) - \sum_{j=3}^{6} I(\mathsf{W}_j; \mathsf{W}_{[1:j-1]})
\end{aligned} \tag{3.62}
$$

where (a) follows from (3.61); (b) follows from the fact that W_i is a function of $(\overline{\mathsf{S}}_{\to i}, \mathsf{W}_{[1:i-1]}) = \mathsf{L}_{i-1,i}$ by Lemma 1; (c) follows from the fact that $\mathsf{S}_{[i+1:j-1]\to i}$ is a function of $\mathsf{W}_{[1:j-1]}$; and (d) follows from the fact that $\mathsf{S}_{j\to[2:j-1]}$ is a function of W_j. Further note that

$$\begin{aligned}
&\sum_{j=3}^{6} H(\mathsf{S}_{j\to[2:j-1]}) + 2\alpha \overset{(a)}{=} \sum_{j=3}^{6} H(\mathsf{S}_{j+1\to[3:j]}) + 2\alpha \\
&= \sum_{j=4}^{7} H(\mathsf{S}_{j\to[3:j-1]}) + 2\alpha \geq H(\overline{\mathsf{S}}_{\to[3:6]}) + 2\alpha \\
&\overset{(b)}{\geq} H(\overline{\mathsf{S}}_{\to[3:6]}) + H(\mathsf{W}_1) + H(\mathsf{W}_2) \\
&\geq H(\overline{\mathsf{S}}_{\to[3:6]}, \mathsf{W}_{[1:2]}) = H(\mathsf{L}_{2,6})
\end{aligned} \tag{3.63}$$

where (a) follows from the fact that $H(\mathsf{S}_{j\to[2:j-1]}) = H(\mathsf{S}_{j+1\to[3:j]})$ due to the symmetrical

code that we consider; and (b) is due to the node-capacity constraint. We thus have

$$
\begin{aligned}
&8\alpha + 3H(\mathsf{S}_{\to 1}) \\
&\overset{(a)}{\geq} 8\alpha + \sum_{i=2}^{6} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\underline{\mathsf{S}}_{\to 4}) \\
&\overset{(b)}{\geq} 8\alpha + \sum_{j=3}^{6} H(\mathsf{S}_{j\to[2:j-1]}) + H(\mathsf{W}_{[1:6]}) - \sum_{j=3}^{6} I(\mathsf{W}_j; \mathsf{W}_{[1:j-1]}) \\
&\overset{(c)}{\geq} 6\alpha + H(\mathsf{L}_{2,6}) + H(\mathsf{W}_{[1:6]}) - \sum_{j=3}^{6} I(\mathsf{W}_j; \mathsf{W}_{[1:j-1]}) \\
&\overset{(d)}{\geq} \sum_{j=1}^{6} H(\mathsf{W}_j) + H(\mathsf{L}_{2,6}) + H(\mathsf{W}_{[1:6]}) - \sum_{j=3}^{6} I(\mathsf{W}_j; \mathsf{W}_{[1:j-1]}) \\
&= \sum_{j=2}^{6} I(\mathsf{W}_j; \mathsf{W}_{[1:j-1]}) + H(\mathsf{L}_{2,6}) + 2H(\mathsf{W}_{[1:6]}) \\
&\qquad - \sum_{j=3}^{6} I(\mathsf{W}_j; \mathsf{W}_{[1:j-1]}) \\
&= I(\mathsf{W}_1; \mathsf{W}_2) + H(\mathsf{L}_{2,6}) + 2H(\mathsf{W}_{[1:6]}) \\
&\geq H(\mathsf{L}_{2,6}) + 2H(\mathsf{W}_{[1:6]}) \overset{(e)}{\geq} 3H(\mathsf{W}_{[1:6]}) \overset{(f)}{=} 3H(\mathsf{W}_{[1:7]}, \mathsf{M}) \\
&\overset{(g)}{=} 3H(\mathsf{W}_{[1:7]}, \mathsf{M}, \mathsf{S}_{\to 1}) \geq 3H(\mathsf{M}, \mathsf{S}_{\to 1}) \\
&= 3H(\mathsf{S}_{\to 1}) + 3H(\mathsf{M}|\mathsf{S}_{\to 1}) \overset{(h)}{=} 3H(\mathsf{S}_{\to 1}) + 3H(\mathsf{M}) \\
&= 3H(\mathsf{S}_{\to 1}) + 3B
\end{aligned}
$$

where (a) follows from (3.57); (b) follows from (3.62); (c) follows from (3.63); (d) follows from the node-capacity constraint; (e) follows from the fact that $H(\mathsf{L}_{2,6}) \geq H(\mathsf{W}_{[1:6]})$ due to Lemma 1; (f) follows from the facts that M is a function of $\mathsf{W}_{[1:6]}$ due to the message-recovery requirement (3.2) and that W_7 is a function of $\mathsf{S}_{\to 7}$, which is in turn a function of $\mathsf{W}_{[1:6]}$; (g) follows from the fact that $\mathsf{S}_{\to 1}$ is a function of $\mathsf{W}_{[2:7]}$; and (h) follows from the repair-secrecy constraint (3.4) with $\ell = 1$. Canceling $3H(\mathsf{S}_{\to 1})$ from both

sides and normalizing the remaining terms by B complete the proof of (3.15) for the cases where $\alpha \geq H(\mathsf{S}_{[2:4]\to 1})$.

4. CODED CACHING

4.1 Introduction

The caching problem is an open problem in network coding. This system is proposed to balance the usage of a network between its peak-traffic time and off-peak time: each user receives messages from the central server and store in local cache in advance during off-peak time in order to reduce the data transmission when the user requests some file from the central server during the peak-traffic time.

More specifically, shown as in Figure 4.1, we have N files and K users in this system. Each user has a local cache with M bits of memory size. This system works in two phases:

- caching phase: This phase is corresponding to the off-peak time of the network. In this phase, without knowing any user's request in the next phase in prior, the central server send messages coded on all N files to each user individually. Thus different users may get different messages in this phase. Each user stores the received message into its local cache memory.

- broadcasting phase: This phase is corresponding to the peak-traffic time of the net-

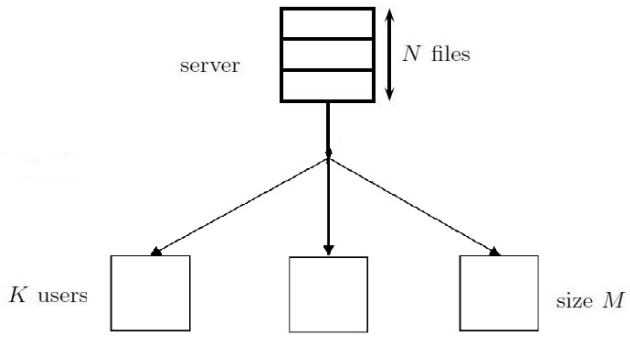


Figure 4.1: The file recovery constraint for $(4, 3, 3)$ secure exact-repair regenerating codes.

work. In this phase, each user requests a file over all N files. By knowing all users' requests, the central server broadcast an encoded message to all users. Thus all users get a same message in this phase. Each user can decode the file requested by itself according to its cache memory content and the broadcast message it received.

Our goal is to characterize the tradeoff region for all possible *memory-capacity* and *broadcast-rate* pairs.

Maddah-Ali and Niesen [15] raised an inner bound for the caching problem in [15]. In their coding scheme, the inner bound can achieve $K + 1$ corner points, thus K linear segments for an (N, K) caching system. Our conjecture is the optimality of this inner bound. The main work in this paper is to prove the optimality of three segments of them.

4.2 Problem Formulation and the Maddah Ali-Niesen Conjecture

For given integer N and K, let W_i, $i = 1, 2, \ldots, N$ be N independent random variables and we have K total users, each of which has a cache. We claim a rate pair (M, R) is achievable for a (N, K) caching system as long as

- for each $k \in [K]$, we have a cache encoding function $f_i : [2^F]^N \leftarrow [2^{\lfloor FM \rfloor}]$;
- for each $(d_1, d_2, \ldots, d_k) \in [N]^K$, we have a message encoding function $g_{(d_1,d_2,\ldots,d_K)} : [2^F]^N \leftarrow [2^{\lfloor FR \rfloor}]$;
- for each $(d_1, d_2, \ldots, d_k), k \in [N]^K \times K$, we have a user decoding function $\mu_{(d_1,d_2,\ldots,d_k),k} : [2^{\lfloor FR \rfloor}] \times [2^{\lfloor FM \rfloor}] \leftarrow [2^F]$;

for some integer F, such that

- W_i uniformly distributed over

$$\mathcal{W} = \{1, 2, 3, \ldots, 2^F\}, \tag{4.1}$$

thus W_i represent a file of F bits;

- encode all files $(W_1, W_2, \ldots, W_N)$ for all K cache, when $k \in [K]$, we have

$$Z_k = f_k(W_1, W_2, \ldots, W_N) \tag{4.2}$$

as the k-th cache content;

- encode all files $(W_1, W_2, \ldots, W_N)$ under users' demand, where $(d_1, d_2, \ldots, d_k) \in [N]^K$ means the user k demand W_{d_k}, we have

$$X_{(d_1,d_2,\ldots,d_k)} = g_{(d_1,d_2,\ldots,d_k)}(W_1, W_2, \ldots, W_N) \tag{4.3}$$

as the encoded message to broadcast to all users;

- for the k-th user, together with its cache content Z_k and received message $X_{(d_1,d_2,\ldots,d_k)}$, it can decode

$$\hat{W}_{(d_1,d_2,\ldots,d_k),k} = \mu_{(d_1,d_2,\ldots,d_k),k}(X_{(d_1,d_2,\ldots,d_k)}, Z_k) \tag{4.4}$$

and $\hat{W}_{(d_1,d_2,\ldots,d_k),k} = W_{d_k}$ with 0 error.

We define $\mathcal{R}_{(N,K)}$ as the collection of all achievable (M, R) rate pair for (N, K) caching system.

In [15], Maddah-Ali and Niesen proposed a coding scheme for (N, K) cache system, characterizing a inner bound for the tradeoff rate region between cache memory size M and broadcasting transmission rate R.

Theorem 6 (Madda-Ali and Niesen Inner Bound)**.** *In a (N, K) cache system, where $N, K \in \mathbb{N}$ and $N \geq K$, for any $M = \frac{N}{K}r$, $r = 0, 1, \ldots, K$, pair $(M, R(M))$ is achievable*

where

$$R(M) = K(1 - \frac{M}{N})\frac{1}{1 + KM/N}. \tag{4.5}$$

This scheme gives $K + 1$ corner points, and hence we can achieve all points lying on the segment bounded by any two corner points by time sharing algorithm, as well as the points above those lines. As a corollary, we have

Corollary 5. *In* (N, K) *cache system, any* memory-rate *pair* (M, R) *satisfying*

$$K(K+1)M + (K+1-r)(K+2-r)NR \geq (K+r)(K+1-r)N \tag{4.6}$$

$\forall r = 1, 2, \ldots, K$ *is achievable.*

We can see that Maddah-Ali and Niesen's coding scheme gives an achievable rate region bounded by K non-trivial linear segments for a (N, K) cache system. For example, Figure.4.2 shows the *memory-rate* tradeoff region for $N = 6$ files and $K = 3$ users. Points $(0, 3)$, $(2, 1)$, $(4, 1/3)$ and $(6, 0)$ are achieved by equation (4.5) with $r = 0, 1, 2, 3$ respectively. Convex hull of the achievable points gives us 3 linear segments.

Naturally, we will question the optimality of this coding scheme. In following parts, I will show that some segments of this scheme is tight, thus the outer bound coincide with the inner bound, when the number of files N is sufficiently large. Further more, with the upcoming results, we can see that the *memory-rate* tradeoff region we presented in Figure4.2 for $(6, 3)$ cache system is tight. More than that, we can characterize the optimal *memory-rate* tradeoff region for general N and $K = 3$.

In order to reduce ambiguity, we give order of segments from the bottom to the top so that the order of segment consist with formula (4.6). That is to say, we will call the segment as the i-th segment if it is the i-th segment counting from bottom to the top,

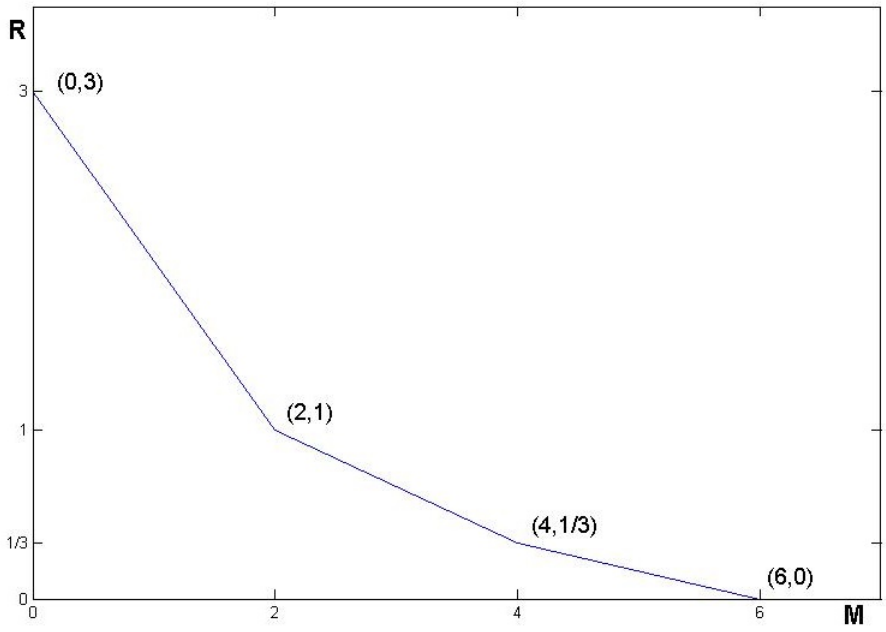


Figure 4.2: *Memory-rate* tradeoff region for $N = 6$ files and $K = 3$ users. Blue curve is obtained by formula (4.6). All area above the blue curve, including the blue curve, is achievable.

moreover the formula of the i-th segment satisfies $r = i$ in the form of (4.6). I will prove the optimality of the first, second and the last segments in following sections respectively in this paper.

4.3 Symmetries

Symmetry under permutation is one of the basic properties of the combinatorial approach. In Chapter 2 and Chapter 3, symmetry property is applied, giving us freedom on operating random variables and hence get the data structure we want. Similarly, the Symmetry property still exists in caching problem, however in caching problem the symmetry is even more complex than what in regenerating codes problems. Formed in two parts, the symmetry argument in caching is discovered and raised by Tian in [21].

4.3.1 Symmetry in Users

Firstly we define a permutation function $\tilde{\pi}$ that

$$\tilde{\pi} : [K] \to [K] \tag{4.7}$$

as a permutation of user index. Correspondingly, the inverse permutation of $\tilde{\pi}$ is defined as $\tilde{\pi}^{-1}(\cdot)$.

With $\tilde{\pi}$ and $\tilde{\pi}^{-1}$, we can define a function $\tilde{\Pi}$ such that

$$\tilde{\Pi}(Z_k) = Z_{\tilde{\pi}(k)}, k \in [K] \tag{4.8}$$

$$\tilde{\Pi}(X_{(d_1,d_2,\ldots,d_K)}) = X_{(d_{\tilde{\pi}^{-1}(1)},d_{\tilde{\pi}^{-1}(2)},\ldots,d_{\tilde{\pi}^{-1}(K)})} \tag{4.9}$$

and for $\mathcal{Z}$ as a collection of Z_k, $\mathcal{X}$ as a collection of X and $\mathcal{W}$ as a collection of W_i, we define $\tilde{\Pi}$ as

$$\tilde{\Pi}(\mathcal{Z}) = \{\tilde{\Pi}(Z_k) : Z_k \in \mathcal{Z}\} \tag{4.10}$$

$$\tilde{\Pi}(\mathcal{X}) = \{\tilde{\Pi}(X) : X \in \mathcal{X}\} \tag{4.11}$$

The symmetry in users means that without losing generality we can assume

$$H(\mathcal{W}, \mathcal{Z}, \mathcal{X}) = H\left(\mathcal{W}, \tilde{\Pi}(\mathcal{Z}), \tilde{\Pi}(\mathcal{X})\right) \tag{4.12}$$

4.3.2 Symmetry in Files

We define a permutation function $\hat{\pi}$ that

$$\hat{\pi} : [N] \to [N] \tag{4.13}$$

as permutation of file index. We still use $\mathcal{Z}$ as a collection of Z_k, $\mathcal{X}$ as a collection of X and $\mathcal{W}$ as a collection of W_i. Based on $\hat{\pi}$, we define a function $\hat{\Pi}$ that

$$\hat{\Pi}(W_i) = W_{\tilde{\pi}(i)}, i \in [N] \tag{4.14}$$

$$\hat{\Pi}(X_{(d_1,d_2,\ldots,d_K)}) = X_{(d_{\tilde{\pi}(1)},d_{\tilde{\pi}(2)},\ldots,d_{\tilde{\pi}(K)})} \tag{4.15}$$

and

$$\hat{\Pi}(\mathcal{W}) = \{\hat{\Pi}(W_i) : W_i \in \mathcal{W}\} \tag{4.16}$$

$$\hat{\Pi}(\mathcal{X}) = \{\hat{\Pi}(X) : X \in \mathcal{X}\} \tag{4.17}$$

The symmetry in files gives us that without losing generality we can assume

$$H(\mathcal{W}, \mathcal{Z}, \mathcal{X}) = H\left(\hat{\Pi}(\mathcal{W}), \mathcal{Z}, \hat{\Pi}(\mathcal{X})\right) \tag{4.18}$$

The detailed proof and argument of symmetry can be found in [21].

4.4 Optimality of the First Segment

Theorem 7 (Outer Bound for the First Segment)**.** *In a* (N, K) *cache system, any achievable* memory-rate *pair* (M, R) *satisfies*

$$M + NR \geq N \tag{4.19}$$

Proof. The proof of this segment is straightforward. For our convenience, we define

$X_{(i,i,\ldots,i)}$, a special type of X, as X_i, for $i = 1, 2, \ldots, N$. Therefore we have

$$
\begin{aligned}
& MF + NRF \\
\geq & H(Z_1) + NH(X_1) && (4.20)\\
= & H(Z_1) + \sum_{i=1}^{N} H(X_i) && (4.21)\\
\geq & H(Z_1, X_1, X_2, \ldots, X_N) && (4.22)\\
\geq & H(Z_1, W_1, W_2, \ldots, W_N) && (4.23)\\
\geq & NF && (4.24)
\end{aligned}
$$

By canceling F from both side of inequality, we have

$$M + NR \geq N \tag{4.25}$$

as our first segment. □

4.5 Optimality of the Second Segment

Before I prove the tightness of this segment, I need to define some notations for our convenience. Let's consider a type of transmitted message X which has a special form of demanding: except the k-th user demanding W_j, every other user demand W_i. That is to say, we focus on the type of message in the form of $X_{(i,\ldots,i,j,i,\ldots,i)}$, for $i, j \in [N]$.

Therefore for this kind of messages, we can distinguish each other by only 3 different things: file index i as the index of file demanded by $K - 1$ users, file index j as the index of file demanded by the special user and the node index k as the index of user who has demand for j.

Hence, we can write all message in the form $X_{(i,\ldots,i,j,i,\ldots,i)}$ of as $X_{i,j,k}$ for short without any ambiguity. Furthermore, we use $X_{j,k}$ short for $X_{1,j,k}$. For a collection of this type of

messages, we use $X_{[j_1:j_2],k}$ to denote $X_{j_1,k}, \ldots, X_{j_2,k}$ for some $j_2 \geq j_1$ and $X_{[j_1:j_2],[k_1:k_2]}$ to denote $X_{[j_1:j_2],k_1}, \ldots, X_{[j_1:j_2],k_2}$ for some $k_2 \geq k_1$.

Also, I need to mention that, for those $X_{j,k}$ and the collection of it that $j \notin [N]$ or $k \notin [K]$, then we regard it as an empty set.

In order to prove the optimality of the second segment, I will show that

Theorem 8 (Outer Bound for the Second Segment)**.** *In a* (N, K) *cache system where* $N > \frac{K+1}{2}$*, any achievable* memory-rate *pair* (M, R) *satisfies*

$$K(K+1)M + K(K-1)NR \geq (K+2)(K-1)N \tag{4.26}$$

The formula (4.26) can be obtained by taking $r = 2$ in inequality (4.6), Corollary 5. That is to say, Theorem 8 is a conversion of Corollary 5 when $r = 2$ particularly, and hence it can prove the optimality of the bound obtained by (4.6) when $r = 2$. In the proving Theorem 8, following lemmas play a very important role.

Lemma 6. *For any* (N, K) *cache code, and with an integer* $p \leq N$*, we have*

$$H(Z_1) + pH(X_{2,K}) \geq H(Z_1, W_{[p]}) + pH(X_{p+1,K}|Z_1, W_{[p]}) \tag{4.27}$$

Lemma 6 says multiple messages can working with one user's cache memory and decode multiple files and the residue. The residue can still work with other users' cache memory, which will be shown in following.

Lemma 7. *For any* (N, K) *cache code that* $K \geq 2$*, we have*

$$\begin{aligned} &H(W_{[p]}, X_{[p+1:N],[3:K]}) + (K-2)H(Z_1, W_{[p]}) \\ &+ \frac{(K-2)(K-3)}{2}(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \\ \geq &(K-2)NF + pF \end{aligned} \tag{4.28}$$

Lemma 8. *For any* $K \geq 2$

$$KH(Z_1, W_{[p]}) + \frac{K(K-1)}{2}(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \geq (K-1)N + p \tag{4.29}$$

Lemma 7 is used in proving Lemma 8, and we need Lemma 8 in proving Theorem 8. These lemmas are proved in Appendix C. After introducing these important lemmas, now we can prove Theorem 8.

Proof. We write $(K-1)N = (K+1)p + q$, where $p, q \in \mathbb{N}$, $q \leq K$. Therefore

$$
\begin{aligned}
& K(K+1)MF + K(K-1)NRF \\
=& K(K+1)MF + K[(K+1)p+q]RF \\
=& (K+1-q)(KMF + KpRF) \\
& + q[KMF + K(p+1)RF] && (4.30) \\
\geq& (K+1-q)[KH(Z_1) + KpH(X_{2,K}) \\
& + q[KH(Z_1) + K(p+1)H(X_{2,K})] && (4.31) \\
\geq& (K+1-q)[KH(Z_1, W_{[p]}) + KpH(X_{p+1,K}|Z_1, W_{[p]}) \\
& + q[KH(Z_1, W_{[p+1]}) + KpH(X_{p+1,K}|Z_1, W_{[p]}) + KH(X_{p+2,K}|Z_1, W_{[p+1]})] && (4.32) \\
=& (K+1-q)KH(Z_1, W_{[p]}) + qKH(Z_1, W_{[p+1]}) \\
& + [K(K+1)pH(X_{p+1,K}|Z_1, W_{[p]}) + qKH(X_{p+2,K}|Z_1, W_{[p+1]})] && (4.33)
\end{aligned}
$$

(4.32) follows Lemma 6. In our definition, p and q are the quotient and remainder of $(K-1)N$ divide $(K+1)$. If we take a close look on the item of $[\cdot]$ (4.33), we have

$$
\begin{aligned}
& K(K+1)pH(X_{p+1,K}|Z_1, W_{[p]}) + qKH(X_{p+2,K}|Z_1, W_{[p+1]}) \\
\geq& (K+1-q)\frac{K(K-1)}{2}(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \\
& + q\frac{K(K-1)}{2}(N-p-1)H(X_{p+2,K}|Z_1, W_{[p+1]}) && (4.34)
\end{aligned}
$$

(4.34) holds as following:

- The summation of the coefficients in both sides are equal as following:

$$
\begin{aligned}
&(K+1-q)\frac{K(K-1)}{2}(N-p)+q\frac{K(K-1)}{2}(N-p-1) \\
=&\frac{K(K-1)}{2}[(K+1-q)(N-p)+q(N-p)-q] && (4.35)\\
=&\frac{K(K-1)}{2}[(K+1)(N-p)-q] && (4.36)\\
=&\frac{K(K-1)}{2}[(K+1)(N-\frac{(K-1)N-q}{K+1})-q] && (4.37)\\
=&\frac{K(K-1)}{2}[(K+1)\frac{2N+q}{K+1}-q] && (4.38)\\
=&\frac{K(K-1)}{2}2N && (4.39)\\
=&K(K-1)N && (4.40)\\
=&K(K+1)p+Kq. && (4.41)
\end{aligned}
$$

- The coefficient of item $H(X_{p+2,K}|Z_1, W_{[p+1]})$ in the LHS is smaller than that in the RHS that:

$$q\frac{K(K-1)}{2}(N-p-1) \geq qK \tag{4.42}$$

(4.42) follows

$$p \leq \frac{K-1}{K+1}N < N-1, \tag{4.43}$$

which can be derive from our definition that $N > \frac{K+1}{2}$ and p is an integer.

- For the items in the both sides, we have

$$H(X_{p+2,K}|Z_1, W_{[p+1]}) \leq H(X_{p+1,K}|Z_1, W_{[p]}) \tag{4.44}$$

holds trivially.

Therefore, we can prove (4.34). Next, substituting (4.34) into (4.33) gives:

$$
\begin{aligned}
&(K+1-q)KH(Z_1, W_{[p]}) + qKH(Z_1, W_{[p+1]}) \\
&\quad + [K(K+1)pH(X_{p+1,K}|Z_1, W_{[p]}) + qKH(X_{p+2,K}|Z_1, W_{[p+1]})] \\
\geq &(K+1-q)[KH(Z_1, W_{[p]}) + \frac{K(K-1)}{2}(N-p)H(X_{p+1,K}|Z_1, W_{[p]})) \\
&\quad + q[KH(Z_1, W_{[p+1]}) + \frac{K(K-1)}{2}(N-p-1)H(X_{p+2,K}|Z_1, W_{[p+1]})] \qquad (4.45) \\
\overset{(a)}{\geq} &(K+1-q)[(K-1)NF + pF] + q[(K-1)NF + (p+1)F] \qquad (4.46) \\
= &(K+2)(K-1)NF, \qquad (4.47)
\end{aligned}
$$

where (a) follows Lemma 8. That is to say,

$$K(K+1)MF + K(K-1)NRF \geq (K+2)(K-1)NF \qquad (4.48)$$

Hence Theorem 8 is proved by canceling F from both sides. □

4.6 Optimality of the Last Segment

In this section, like what we do in last section, we also focus on some special type of user demand and corresponding transmitted message. When $N \geq \frac{K(K+1)}{2}$, We denote message $X_{(j,k)}$ (distinguish from $X_{j,k}$ with extra $(\cdot)$ in the index) as

$$X_{(j,k)} = X_{(1,2,\ldots,k-1,j,k+1+(N-K),\ldots,N)}, \qquad (4.49)$$

thus the first $k-1$ users demand files with index 1 to $k-1$, the last $K-k$ users demand files with index $k+1+(N-K)$ to N, and the k-th user demand file with index j. For physically meaningful, we have $k \leq j \leq k+(N-K)$, else we regard it as $\emptyset$. We have

following outer bound:

Theorem 9 (Outer Bound for the Last Segment)**.** *In a (N, K) cache system, for any $N \geq \frac{K(K+1)}{2}$, all achievable (M, R) pairs satisfy:*

$$K(K+1)M + 2NR \geq 2NK \tag{4.50}$$

Following lemmas are essential in our proof:

Lemma 9. *For $k = 0, 1, \ldots, K-1$, $N \geq \frac{K(K+1)}{2}$ and $i \in \{k, k+1, \ldots, N-(K-k)\}$, $1 \leq A \leq N-(K-k)+1-i$, we have*

$$\begin{aligned} &AH(X_{(i+1,k+1)}|Z_{[k]}, W_{[i]}) + H(Z_{[k+1]}, W_{[i]}) \\ \geq &AH(X_{(i+A+1,k+2)}|Z_{[k+1]}, W_{[i+A]}) + H(Z_{[k+1]}, W_{[i+A]}). \end{aligned} \tag{4.51}$$

Lemma 10. *For a (N, K) caching code, $N \geq \frac{K(K+1)}{2}$, $K \geq 2$, all achievable (M, R) pair satisfies*

$$\begin{aligned} &(K+1)!MF + (K-1)!2NRF \\ \geq &\frac{(K-1)!}{k}[(K(K+1)-q_k)H(Z_{[k]}, W_{[p_k]}) + q_k H(Z_{[k]}, W_{[p_k+1]})] \\ &+ \frac{(K-1)!}{k}[(2kN-q_k)H(X_{(p_k+1,k+1)}|Z_{[k]}, W_{[p_k]}) \\ &\quad + q_k H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]})] \\ &+ (k-1)(K-1)!NF \end{aligned} \tag{4.52}$$

$\forall k \leq K-1$, where integer p_k and q_k satisfy

$$k(k+1)N = K(K+1)p_k + q_k \tag{4.53}$$

and $q_k < K(K+1)$.

Proof is organized in the following way: Lemma 9 is used to prove Lemma 10, and Lemma 10 is the key to prove Theorem 9 . All detailed proofs are in appendix. Now we can proceed to prove (4.50).

Proof. Firstly, Lemma 10 gives:

$$\begin{aligned}
&(K+1)!MF + (K-1)!2NRF \\
\geq &\frac{(K-1)!}{K-1}[(K(K+1)-q_{K-1})H(Z_{[K-1]}, W_{[p_{K-1}]}) + q_{K-1}H(Z_{[K-1]}, W_{[p_{K-1}+1]})] \\
&+ \frac{(K-1)!}{K-1}[(2(K-1)N - q_{K-1})H(X_{(p_{K-1}+1,K)}|Z_{[K-1]}, W_{[p_{K-1}]}) \\
&\quad + q_{K-1}H(X_{(p_{K-1}+2,K)}|Z_{[K-1]}, W_{[p_{K-1}+1]})] \\
&+ (K-2)(K-1)!NF
\end{aligned} \tag{4.54}$$

by taking $k = K-1$. For any $p \leq N$, by Han's Inequality we have

$$H(Z_{[K-1]}, W_{[p]}) \geq \frac{K-1}{K}H(Z_{[K]}, W_{[p]}) + \frac{1}{K}H(W_{[p]}) \tag{4.55}$$

Substituting into (4.54) implies:

$$\begin{aligned}
&(K+1)!MF + (K-1)!2NRF \\
\geq &\frac{(K-1)!}{K}[(K(K+1)-q_{K-1})H(Z_{[K]}, W_{[p_{K-1}]}) + q_{K-1}H(Z_{[K]}, W_{[p_{K-1}+1]})] \\
&+ \frac{(K-1)!}{K-1}\frac{1}{K}[(K(K+1)-q_{K-1})H(W_{[p_{K-1}]}) + q_{K-1}H(W_{[p_{K-1}+1]})] \\
&+ \frac{(K-1)!}{K-1}[(2(K-1)N - q_{K-1})H(X_{(p_{K-1}+1,K)}|Z_{[K-1]}, W_{[p_{K-1}]}) \\
&\quad + q_{K-1}H(X_{(p_{K-1}+2,K)}|Z_{[K-1]}, W_{[p_{K-1}+1]})] \\
&+ (K-2)(K-1)!NF
\end{aligned} \tag{4.56}$$

In (4.56), since item $H(W_{[p_{K-1}]}) \geq p_{K-1}F$ and $H(W_{[p_{K-1}+1]}) \geq (p_{K-1}+1)F$, therefore we can simplify them as

$$\begin{aligned}
&\frac{(K-1)!}{K-1}\frac{1}{K}[(K(K+1)-q_{K-1})H(W_{[p_{K-1}]}) + q_{K-1}H(W_{[p_{K-1}+1]})] \\
\geq &\frac{(K-1)!}{K-1}\frac{1}{K}[(K(K+1)-q_{K-1})p_{K-1} + q_{K-1}(p_{K-1}+1)] \\
= &\frac{(K-1)!}{K-1}\frac{1}{K}[K(K+1)p_{K-1}F + q_{K-1}F] \\
= &\frac{(K-1)!}{K-1}\frac{1}{K}K(K-1)NF \\
= &(K-1)!NF
\end{aligned} \tag{4.57}$$

Further more, we can lower bound some items in (4.56) as:

$$\begin{aligned}
&\frac{(K-1)!}{K-1}[(2(K-1)N - q_{K-1})H(X_{(p_{K-1}+1,K)}|Z_{[K-1]}, W_{[p_{K-1}]}) \\
&\qquad + q_{K-1}H(X_{(p_{K-1}+2,K)}|Z_{[K-1]}, W_{[p_{K-1}+1]})] \\
\geq &\frac{(K-1)!}{K}[(K(K+1)-q_{K-1})H(X_{(p_{K-1}+1,K)}|Z_{[K-1]}, W_{[p_{K-1}]}) \\
&+ K(K+1)(p_K - p_{K-1} - 1)H(X_{(p_{K-1}+2,K)}|Z_{[K-1]}, W_{[p_{K-1}+1]})]
\end{aligned} \tag{4.58}$$

(4.58) holds according to following:

- The summation of the coefficients of both sides are equal, which is derived as following:

$$\begin{aligned}
&K(K+1) - q_{K-1} + K(K+1)(p_K - p_{K-1} - 1) \\
= &K(K+1) - q_{K-1} + K(K+1)N - K(K+1)p_{K-1} - K(K+1) \\
= &2KN \\
= &\frac{K}{K-1}[(2(K-1)N - q_{K-1}) + q_{K-1}].
\end{aligned} \tag{4.59}$$

- The coefficient of item $H(X_{(p_{K-1}+1,K)}|Z_{[K-1]}, W_{[p_{K-1}]})$ in the LHS is larger than that in the RHS, which follows:

$$q_{K-1} \leq K(K+1) \leq 2N \leq 2N(K-1)^2 \tag{4.60}$$

as our definition. Hence, equivalently we have:

$$\frac{1}{K-1}(2(K-1)N - q_{K-1}) \geq \frac{1}{K}(K(K+1) - q_{K-1}) \tag{4.61}$$

- Trivially we have

$$H(X_{(p_{K-1}+1,K)}|Z_{[K-1]}, W_{[p_{K-1}]}) \geq H(X_{(p_{K-1}+2,K)}|Z_{[K-1]}, W_{[p_{K-1}+1]}) \tag{4.62}$$

Therefore we have (4.57) proved. If we substitute (4.57) and (4.58) into (4.56), we have:

$$\begin{aligned}
&(K+1)!MF + (K-1)!2NRF \\
\geq &\frac{(K-1)!}{K}[(K(K+1) - q_{K-1})H(Z_{[K]}, W_{[p_{K-1}]}) + q_{K-1}H(Z_{[K]}, W_{[p_{K-1}+1]})] \\
&+ \frac{(K-1)!}{K}[(K(K+1) - q_{K-1})H(X_{(p_{K-1}+1,K)}|Z_{[K-1]}, W_{[p_{K-1}]}) \\
&+ K(K+1)(p_K - p_{K-1} - 1)H(X_{(p_{K-1}+2,K)}|Z_{[K-1]}, W_{[p_{K-1}+1]})] \\
&+ (K-2)(K-1)!NF + (K-1)!NF && (4.63) \\
\overset{(a)}{\geq} &\frac{(K-1)!}{K}K(K+1)H(Z_{[K]}, W_{[p_{K-1}]}) + (K-1)(K-1)!NF && (4.64) \\
\geq &(K-1)!(K+1)NF + (K-1)!(K-1)NF && (4.65) \\
= &(K-1)!2NKF && (4.66)
\end{aligned}$$

where (a) follows Lemma 9 by taking $A = p_K - p_{K-1}$, $i = p_{K-1}$ and $A = p_K - p_{K-1} - 1$, $i = p_{K-1} + 1$ respectively.

By canceling F and $(K-1)!$ from both sides of (4.66), we have:

$$K(K+1)M + 2NR \geq 2NK \tag{4.67}$$

Hence our Theorem 9 is proved. □

5. SUMMARY AND CONCLUSIONS

- In this paper, we consider the (n, d) multilevel diversity coding with regeneration problem, which is proposed in [16] firstly. We proved the MBR point can be achieved by separate coding scheme. This proof includes proving two outer bounds. The horizontal one is optimal and another one is not, and they intersect on the MBR point.

  There are several directions to proceed the research. The first is trying to get an optimal outer bound for the segment left to the MBR point. The second is to understand why mixing content can beat over the separate coding and how much it can do.

- In this paper, we considered the (n, k, d, ℓ) secure exact-repair regenerating code problem, which has been previously studied in [13, 14, 17–19]. We proved that when the secrecy parameter ℓ is sufficiently large, the SRK point [19] is the *only* corner point of the achievable normalized storage-capacity repair-bandwidth tradeoff region. This includes all previous results from [13] and [14] as special cases. On the other hand, when ℓ is small, we showed that it is entirely possible that the achievable normalized storage-capacity repair-bandwidth tradeoff region features *multiple* corner points. In particular, we showed that the achievable normalized storage-capacity repair-bandwidth tradeoff region for the $(7, 6, 6, 1)$ problem has exactly *two* corner points. This suggests a much "smoother" transition, in terms of the rate region, from the original exact-repair regenerating code problem to the secrecy extension than that suggested by the previous results from [13] and [14].

  The question whether (3.8) is also *necessary* for the SRK point [19] to be the *only* corner point of the achievable normalized storage-capacity repair-bandwidth trade-

off region remains open. Significant research is also needed to further understand how the tradeoff region $\mathcal{R}_{n,k,d,\ell}$ may look like when ℓ is small (the non-secrecy case with $\ell = 0$ remains open and appears to be very challenging). In particular, one may consider generalizing the code construction given in Section 3.4, which was based on the so-called *canonical* layered codes proposed in [20] for the case of $k = d = n - 1$. The canonical layered codes have been generalized to the cases where $d < n - 1$ and $k < d$ in [20], and further improvements can also be found in [22]. Following the approach used in Section 3.4, the code constructions from [20] and [22] can be similarly adapted into constructions that satisfy the repair-secrecy requirement. The optimalities of these code constructions are currently under our investigations.

- In this paper we consider (N, K) coded caching problems. In [19], a conjecture was proposed without proof. They think the optimal tradeoff region of an (N, K) coded caching problem can be characterized by K linear segments. We proved the first, the second and the last segments, counting from the bottom to the top.

  The direction of future research is to prove the optimality of the segments in between the second and last. To prove the third segment is a good starting point, and it is also very important to find a combinatorial structure for this problem and to find the physical meaning of which.

- Generally, in this paper we proposed a new approach on obtaining outer bounds for variant network coding problems, and showed the effectiveness of this approach. In many network coding problems, there's a symmetry in the system, which makes combinatorial approach possible. We showed how to find the physical meaning for such symmetry we mentioned, hence to derive the combinatorial structure for each problem specifically and apply our combinatorial approach.

The directions of future research on combinatorial approach are following:

- applying combinatorial approach to more networking problems, getting the combinatorial structure of which and understand the physical meaning of those structures eventually and gradually.
- using combinatorial structure and argument to help computational approach lowing the number of constraints, hence making computational approach practically available in more cases with larger parameter.

REFERENCES

[1] R. C. Singleton, “Maximum distance q-nary codes," *IEEE Trans. Inf. Theory*, vol. IT-10, pp. 116–118, Apr. 1964.

[2] J. R. Roche, “Distributed information storage," *Ph.D. Dissertation*, Stanford University, Stanford, CA, USA, Mar. 1992.

[3] J. R. Roche, R. W. Yeung, and K. P. Hau, “Symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1059–1064, May 1997.

[4] R. W. Yeung and Z. Zhang, “On symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 45, pp. 609–621, Mar. 1999.

[5] S. Mohajer, C. Tian, and S. N. Diggavi, “Asymmetric multilevel diversity coding and asymmetric Gaussian multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4367–4387, Sep. 2010.

[6] J. Jiang, N. Marukala, and T. Liu, “Symmetrical multilevel diversity coding and subset entropy inequalities," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 84–103, Jan. 2014.

[7] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

[8] C. Tian, “Characterizing the rate region of the $(4, 3, 3)$ exact-repair regenerating codes," *IEEE J. Sel. Are. Communications*, vol. 32, no. 5, pp. 967–975, May 2014.

[9] K. V. Rashmi, N. B. Shah, and P. V. Kumar, “Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5227–5239, Aug. 2011.

[10] Cover, Thomas M. and Joy A, *Thomas Elements of information theory*. John Wiley & Sons, 2012.

[11] I. M. Duursma, “Outer bounds for exact repair codes," *Preprint*. [Online] `http://arxiv.org/abs/1406.4852`

[12] S. Shao, T. Liu, and C. Tian, “Multilevel diversity coding with regeneration: Separate coding achieves the MBR point," in *Proc. 50th Ann. Conf. Inf. Sci. Systems (CISS)*, Princeton, NJ, USA, Mar. 2016, pp. 602–607.

[13] R. Tandon, S. Amuru, T. C. Clancy, and R. M. Buehrer, “Towards optimal secure distributed storage systems with exact repair," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3477–3492, Jun. 2016.

[14] F. Ye, K. W. Shum, and R. W. Yeung, “The rate region of secure exact-repair regenerating codes for 5 nodes," in *Proc. IEEE Int. Sym. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1406–1410.

[15] M. A. Maddah-Ali, U. Niesen,"Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May. 2014.

[16] C. Tian and T. Liu, “Multilevel diversity coding with regeneration," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4833–4847, June. 2016.

[17] S. Goparaju, S. El Rouayheb, R. Calderbank, and H. V. Poor, “Data secrecy in distributed storage systems under exact repair," in *Proc. IEEE Int. Sym. Net. Coding (NetCod)*, Calgary, AB, Canada, Jun. 2013, pp. 1–6.

[18] S. Pawar, S. El Rouayheb, and K. Ramchandran, “On secure distributed data storage under repair dynamics," in *Proc. IEEE Int. Sym. Inf. Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 2543–2547.

[19] N. B. Shah, K. V. Rashmi, and P. V. Kumar, “Information-theoretically secure regenerating codes for distributed storage," in *Proc. IEEE Glo. Tel. Conference (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–5.

[20] C. Tian, B. Sasidharan, V. Aggarwal, V. A. Vaishampayan, and P. V. Kumar, “Layered exact-repair regenerating codes via embedded error correction and block designs,” *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1933–1947, Mar. 2015.

[21] C. Tian,"Symmetry, demand types and outer bounds in caching systems",in *Proc. IEEE Int. Sym. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 825–829

[22] S. Goparaju, S. El Rouayheb, and R. Calderbank, “New codes and inner bounds for exact repair in distributed storage systems,” in *Proc. IEEE Int. Sym. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun.–Jul. 2014, pp. 1036–1040.

[23] T. S. Han, “Nonnegative entropy measures of multivariate symmetric correlations," *Inf. Control*, vol. 36, no. 2, pp. 133–156, Feb. 1978.

APPENDIX A

PROOF OF LEMMAS FOR MULTILEVEL DIVERSITY CODING WITH REGENERATION PROBLEM

A.1 Proof of Proposition 1

Proof of Proposition 1. Our main tool for proving Proposition 1 is the following "exchange" lemma between $\mathsf{L}_0^{(j)}$ and $\mathsf{L}_0^{(k)}$ for $j \leq k$.

Lemma 11 (Exchange lemma between $\mathsf{L}_0^{(j)}$ and $\mathsf{L}_0^{(k)}$)**.** *For any symmetrical* $(n = d + 1, d, (N_1, \ldots, N_d), T, S)$ *multilevel diversity regenerating code that satisfies the repair requirement* (2.4)*, we have*

$$\begin{aligned} &\frac{d+1-j}{d-k} H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\ &\quad \geq \frac{d+1-j}{d-k} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \end{aligned} \tag{A.1}$$

for any $k \in [1 : d-1]$ *and* $j \in [1 : k]$*. Here, we use the convention that* $\mathsf{L}_0^{(0)} := \emptyset$*.*

A proof of the lemma can be found in following section. Using this lemma, Proposition 1 can be proved as follows. Fix $k \in [1 : d-1]$, add the inequalities (A.1) for $j \in [1 : k]$, and cancel the common term $\sum_{j=1}^{k-1} H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)})$ on both sides. We have

$$\begin{aligned} &\frac{T_{d,k}}{d-k} H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) \\ &\quad \geq \frac{T_{d,k}}{d-k} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(0)}|\mathsf{M}^{(k)}), \end{aligned}$$

which can be equivalently written as

$$\frac{T_{d,k+1}}{d-k}H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) \geq \frac{T_{d,k}}{d-k}H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}), \tag{A.2}$$

by the facts that $H(\mathsf{L}_0^{(0)}|\mathsf{M}^{(k)}) = 0$ and $T_{d,k} + (d-k) = T_{d,k+1}$. Multiplying both sides of (A.2) by $d-k$ and rearranging the terms complete the proof of (2.16).

A.2 Proof of Proposition 2

Proof of Proposition 2. Our main tool for proving Proposition 2 is the following "exchange" lemma between $\mathsf{L}_0^{(j)}$ and $\mathsf{L}_1^{(k)}$ for $j \leq k$.

Lemma 12 (Exchange lemma between $\mathsf{L}_0^{(j)}$ and $\mathsf{L}_1^{(k)}$)**.** *For any symmetrical* $(n = d+1, d, (N_1, \ldots, N_d), T, S)$ *multilevel diversity regenerating code that satisfies the repair requirement* (2.4)*, we have*

$$\begin{aligned}&\frac{d+1-j}{d-k}H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\ &\quad\geq \frac{d+1-j}{d-k}H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)})\end{aligned} \tag{A.3}$$

for any $k \in [1 : d-1]$ *and* $j \in [1 : k]$*. Here, we again follow the convention that* $\mathsf{L}_0^{(0)} := \emptyset$*.*

A proof of the lemma can be found in the following section. Using this lemma, Proposition 2 can be proved as follows. Fix $k \in [1 : d-1]$, add the inequalities (A.3) for $j \in [1 : k]$, and cancel the common term $\sum_{j=1}^{k-1} H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)})$ on both sides. We have

$$\begin{aligned}&\frac{T_{d,k}}{d-k}H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) \\ &\quad\geq \frac{T_{d,k}}{d-k}H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(0)}|\mathsf{M}^{(k)}),\end{aligned}$$

which is equivalent to (2.19) by the fact that $H(\mathsf{L}_0^{(0)}|\mathsf{M}^{(k)}) = 0$. This completes the proof

of (2.19).

A.3 Proof of Lemma 11

Proof of Lemma 11. Fix $k \in [1 : d-1]$ and $j \in [1 : k]$. Since $j \leq k$ by the assumption, we have $d+1-j > d-k$. Thus, we may write $d+1-j = i(d-k)+p$ for some integer $i \geq 1$ and $p \in [1 : d-k]$. For any $q \in [1 : i-1]$, let

$$\mathcal{B}_q := [j+p+(q-1)(d-k) : j+p+q(d-k)-1].$$

Furthermore, let $\mathcal{B}_0 := [j : j+p-1]$. Then we have $[j : k] = \cup_{q=0}^{i-1} \mathcal{B}_q$. Next, let us show by induction that

$$\begin{aligned} &qH(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\ &\quad \geq qH(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\cup_{r=0}^{i-q}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \end{aligned} \tag{A.4}$$

for any $q \in [1 : i]$.

To prove the base case of $q = 1$, note that

$$\begin{aligned} H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) &= H(\mathsf{L}_j, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\ &= H(\mathsf{S}_{[j+1:d+1],j}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\ &\overset{(a)}{=} H(\mathsf{S}_{[j:k],k+1}, \mathsf{S}_{[k+2:d+1],k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\ &= H(\mathsf{S}_{[j:k],k+1}, \mathsf{L}_{k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}), \end{aligned}$$

where (a) follows by swapping j with $k+1$ for the storage-node indices and the fact that the collection of random variables from $\mathsf{L}_0^{(j-1)}$ is invariant under such a swap. Further note that $\mathsf{S}_{[j:k],k+1}$ is a function of $\{\mathsf{W}_s : s \in [j : k]\}$, which is in turn a function of $\mathsf{L}_0^{(k)}$. It

follows that

$$
\begin{aligned}
&H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\
&= H(\mathsf{L}_0^{(k)}, \mathsf{S}_{[j:k],k+1}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{[j:k],k+1}, \mathsf{L}_{k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\
&\overset{(a)}{\geq} H(\mathsf{L}_0^{(k)}, \mathsf{S}_{[j:k],k+1}, \mathsf{L}_{k+1}|\mathsf{M}^{(k)}) \\
&\qquad + H(\mathsf{S}_{[j:k],k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \qquad \text{(A.5)} \\
&\overset{(b)}{=} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{[j:k],k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\
&= H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\cup_{r=0}^{i-1}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}), \qquad \text{(A.6)}
\end{aligned}
$$

where (a) follows from the submodularity of entropy, and (b) follows again from the fact that $\mathsf{S}_{[j:k],k+1}$ is a function of $\mathsf{L}_0^{(k)}$. This completes the proof of the base case of $q = 1$.

Next, assuming (A.4) holds for some $q \in [i-1]$, we have

$$
\begin{aligned}
&(q+1)H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \geq qH(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) \\
&\qquad + H(\mathsf{S}_{\cup_{r=0}^{i-q}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}). \qquad \text{(A.7)}
\end{aligned}
$$

Consider a one-to-one swapping between the elements of $\mathcal{B}_{i-q}$ and $[k+2 : d+1]$ for the storage-node indices, and note that the collection of random variables from $\mathsf{L}_0^{(j-1)}$ is invariant under such swaps. We can write

$$
\begin{aligned}
&H(\mathsf{S}_{\cup_{r=0}^{i-q}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\
&\quad = H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{S}_{[k+2:d+1],k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\
&\quad = H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_{k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&H(\mathsf{S}_{\cup_{r=0}^{i-q}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) \\
&= H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_{k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) \\
&\overset{(a)}{=} H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_{k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{L}_0^{(k)}, \mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}|\mathsf{M}^{(k)}) \\
&\overset{(b)}{\geq} H(S_{\bigcup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_{k+1}, \mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\
&\overset{(c)}{=} H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\bigcup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}),
\end{aligned} \tag{A.8}
$$

where (a) and (c) are due to the fact that $\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}$ is a function of $\mathsf{L}_0^{(k)}$, and (b) follows from the submodularity of entropy. Substituting (A.8) into (A.7) gives

$$
\begin{aligned}
&(q+1)H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\
&\quad \geq qH(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + \\
&\qquad \Big(H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)})\Big) \\
&\quad = (q+1)H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + \\
&\qquad H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}),
\end{aligned}
$$

which completes the induction step and hence the proof of (A.4).

Set $q = i$ in (A.4). We have

$$\begin{aligned} & iH(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\ & \geq iH(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\mathcal{B}_0,k+1}, \mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) \\ & = iH(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) + \\ & \quad H(\mathsf{S}_{\mathcal{B}_0,k+1}|\mathsf{L}_0^{(j-1)}, \mathsf{M}^{(k)}) \end{aligned} \tag{A.9}$$

where the last equality follows from the chain rule for conditional entropy. Consider a one-to-one swapping between the elements of $\mathcal{B}_0 = [j : j+p-1]$ and $\mathcal{B} := [k+2 : k+p+1]$ for the storage-node indices, and note the collection of random variables $\mathsf{L}_0^{(j-1)}$ is invariant under such swaps. We can write

$$H(\mathcal{S}_{\mathcal{B}_0,k+1}|\mathsf{L}_0^{(j-1)}, \mathsf{M}^{(k)}) = H(\mathsf{S}_{\mathcal{B},k+1}|\mathsf{L}_0^{(j-1)}, \mathsf{M}^{(k)}). \tag{A.10}$$

Now consider two nonempty subsets $\mathcal{B}'$ and $\mathcal{B}''$ of $[k+2 : d+1]$ of the same cardinality p. Consider a permutation π on the storage-node indices such that: 1) only the indices in $[k+2 : d+1]$ are permuted; and 2) $\mathcal{B}'$ are mapped to $\mathcal{B}''$. Note that the collection of random variables from $\mathsf{L}^{(j-1)}$ is invariant under such a permutation. Then by the symmetry of the code, we have $H(\mathsf{S}_{\mathcal{B}',k+1}|\mathsf{L}_0^{(j-1)}, \mathsf{M}^{(k)}) = H(\mathsf{S}_{\mathcal{B}'',k+1}|\mathsf{L}_0^{(j-1)}, \mathsf{M}^{(k)})$. It thus follows from the well-known Han's inequality [23] that

$$\frac{1}{p} H(\mathsf{S}_{\mathcal{B},k+1}|\mathsf{L}_0^{(j-1)}, \mathsf{M}^{(k)}) \geq \frac{1}{d-k} H(\mathsf{L}_{k+1}|\mathsf{L}_0^{(j-1)}, \mathsf{M}^{(k)}). \tag{A.11}$$

Substituting (A.11) into (A.10) gives:

$$\begin{aligned}
&H(\mathcal{S}_{\mathcal{B}_0,k+1}|\mathsf{L}_0^{(j-1)},\mathsf{M}^{(k)})\\
&\quad\geq \frac{p}{d-k}H(\mathsf{L}_{k+1}|\mathsf{L}_0^{(j-1)},\mathsf{M}^{(k)})\\
&\quad\overset{(a)}{\geq} \frac{p}{d-k}H(\mathsf{L}_{k+1}|\mathsf{L}_0^{(k)},\mathsf{M}^{(k)})\\
&\quad= \frac{p}{d-k}\left[H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)})-H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)})\right], \qquad \text{(A.12)}
\end{aligned}$$

where (a) follows from the fact that conditioning reduces entropy. Substituting (A.12) into (A.9) gives:

$$\begin{aligned}
&\left(i+\frac{p}{d-k}\right)H(\mathsf{L}_0^{(k)}|\mathsf{M}^{(k)})+H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)})\\
&\geq \left(i+\frac{p}{d-k}\right)H(\mathsf{L}_0^{(k+1)}|\mathsf{M}^{(k)})+H(\mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}),
\end{aligned}$$

which is equivalent to (A.1) by noting that

$$i+\frac{p}{d-k}=\frac{i(d-k)+p}{d-k}=\frac{d+1-j}{d-k}.$$

This completes the proof of Lemma 11.

A.4 Proof of Lemma 12

Proof of Lemma 12. Fix $k\in[1:d-1]$ and $j\in[1:k]$. Since $j\leq k$ by the assumption, we have $d+1-j>d-k$. Thus, we may write $d+1-j=i(d-k)+p$ for some integer $i\geq 1$ and $p\in[1:d-k]$. For any $q\in[1:i-1]$, let

$$\mathcal{B}_q:=[j+p+(q-1)(d-k):j+p+q(d-k)-1].$$

Furthermore, let $\mathcal{B}_0:=\{1\}\cup[j+1:j+p-1]$. Then we have $\{1\}\cup[j+1:k]=\cup_{q=0}^{i-1}\mathcal{B}_q$.

Next, let us show by induction that

$$\begin{aligned} & qH(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\ & \geq qH(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\cup_{r=0}^{i-q}\mathcal{B}_r,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \end{aligned} \tag{A.13}$$

for any $q \in [1:i]$, where $\mathsf{L}'_{[2:j]} := \{\mathsf{L}_{[2:j]}, \mathsf{S}_{1,[2:j]}\}$.

To prove the base case of $q = 1$, note that

$$\begin{aligned} H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) &\overset{(a)}{=} H(\mathsf{S}_{1,[2:j+1]}, \mathsf{L}_{[2:j+1]}|\mathsf{M}^{(k)}) \\ &= H(\mathsf{S}_{1,j+1}, \mathsf{L}_{j+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\ &= H(\mathsf{S}_{1,j+1}, \mathsf{S}_{[j+2:d+1],j+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\ &\overset{(b)}{=} H(\mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{S}_{[k+2:d+1],k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\ &= H(\mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{L}_{k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}), \end{aligned}$$

where (a) follows by swapping r with $r+1$ for all $r \in [1:d]$ and $d+1$ with 1 for the storage-node indices, and (b) follows by swapping $j+1$ with $k+1$ for the storage-node

indices. We thus have

$$
\begin{aligned}
&H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\
&= H(\mathsf{W}_1, \mathsf{L}_{[2:k]}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{L}_{k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\overset{(a)}{=} H(\mathsf{W}_1, \mathsf{L}'_{[2:k]}, \mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{L}_{k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\overset{(b)}{\geq} H(\mathsf{W}_1, \mathsf{L}'_{[2:k]}, \mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{L}_{k+1}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\overset{(c)}{=} H(\mathsf{W}_1, \mathsf{L}_{[2:k+1]}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{S}_{1,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&= H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\cup_{r=0}^{i-1}\mathcal{B}_r,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}),
\end{aligned} \tag{A.14}
$$

where (a) and (c) follow from the facts that $\mathsf{S}_{1,[2:k+1]}$ is a function of W_1 and that $\mathsf{S}_{[j+1:k],k+1}$ is a function of $\{\mathsf{W}_s : s \in [j+1:k]\}$, which is in turn a function of $\{\mathsf{W}_1, \mathsf{L}_{[2:k]}\}$, and (b) is due to the submodularity of entropy. This completes the proof of the base case of $q = 1$.

Assume that (A.13) holds for some $q \in [1 : i-1]$. We have

$$
\begin{aligned}
&(q+1)H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\
&\geq qH(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(S_{\cup_{r=0}^{i-q}\mathcal{B}_r,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}).
\end{aligned} \tag{A.15}
$$

Consider a one-to-one swapping between the elements of $\mathcal{B}_{i-q}$ and $[k+2 : d+1]$ for the storage-node indices, and note that the collection of random variables from $\mathsf{L}'_{[2:j]}$ is

invariant under such swaps. We have

$$
\begin{aligned}
&H(\mathsf{S}_{\cup_{r=0}^{i-q}\mathcal{B}_r,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) + H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) \\
&= H(S_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{S}_{[k+2:n],k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\quad + H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) \\
&= H(S_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_{k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\quad + H(\mathsf{W}_1, \mathsf{L}_{[2:k]}|\mathsf{M}^{(k)}) \\
&\overset{(a)}{=} H(S_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}_{k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\quad + H(\mathsf{W}_1, \mathsf{L}'_{[2:k]}, \mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}|\mathsf{M}^{(k)}) \\
&\overset{(b)}{\geq} H(\mathsf{W}_1, \mathsf{L}'_{[2:k]}, \mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{S}_{[j+1:k],k+1}, \mathsf{L}_{k+1}|\mathsf{M}^{(k)}) \\
&\quad + H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\overset{(c)}{=} H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}),
\end{aligned} \tag{A.16}
$$

where (a) and (c) are true because $\mathsf{S}_{1,[2;k+1]}$ is a function of W_1 and that $\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}$ is a function of $\{\mathsf{W}_s : s \in \cup_{r=0}^{i-(q+1)}\mathcal{B}_r\}$, which is in turn a function of $\{\mathsf{W}_1, \mathsf{L}_{[2:k]}\}$, and (b) is due to the submodularity of entropy. Substituting (A.16) into (A.15) gives

$$
\begin{aligned}
&(q+1)H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\
&\geq (q+1)H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\cup_{r=0}^{i-(q+1)}\mathcal{B}_r,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}),
\end{aligned}
$$

which completes the induction step and hence the proof of (A.13).

Set $q = i$ in (A.13). We have

$$\begin{aligned}
&iH(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)}) \\
&\geq iH(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{S}_{\mathcal{B}_0,k+1}, \mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) \\
&\overset{(a)}{=} iH(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}'_{[2:j]}|\mathsf{M}^{(k)}) + \\
&\quad H(\mathsf{S}_{\mathcal{B}_0,k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)}) \\
&\overset{(b)}{=} iH(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}) + \\
&\quad H(S_{\mathcal{B}_0,k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)}),
\end{aligned} \tag{A.17}$$

where (a) follows from the chain rule for conditional entropy, and (b) follows by swapping r with $r+1$ for $r \in [1 : d]$ and $d+1$ with 1 for the storage-node indices. Consider a one-to-one swapping between the elements of $\mathcal{B}_0 = \{1\} \bigcup [j+1 : j+p-1]$ and $\mathcal{B} := [k+2 : k+p+1]$ for the storage-node indices, and note the collection of random variables from $\mathsf{L}'_{[2:j]}$ is invariant under such swaps. We can write

$$H(S_{\tau_0,k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)}) = H(S_{\mathcal{B},k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)}) \tag{A.18}$$

Now consider two nonempty subsets $\mathcal{B}'$ and $\mathcal{B}''$ of $[k+2 : d+1]$ of the same cardinality p. Consider a permutation π on the storage-node indices such that: 1) only the indices in $[k+2 : d+1]$ are permuted; and 2) $\mathcal{B}'$ are mapped to $\mathcal{B}''$. Note that the collection of random variables from $\mathsf{L}'_{[2:j]}$ is invariant under such a permutation. Then by the symmetry of the code, we have $H(\mathsf{S}_{\mathcal{B}',k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)}) = H(\mathsf{S}_{\mathcal{B}'',k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)})$. It thus follows from the well-known Han's inequality [23] that

$$\frac{1}{p} H(S_{\mathcal{B},k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)}) \geq \frac{1}{d-k} H(\mathsf{L}_{k+1}|\mathsf{L}'_{[2:j]}, \mathsf{M}^{(k)}). \tag{A.19}$$

Substituting (A.19) into (A.18) gives:

$$\begin{aligned}
&H(S_{\tau_0,k+1}|\mathsf{L}'_{[2:j]},\mathsf{M}^{(k)})\\
&\geq \frac{p}{d-k}H(\mathsf{L}_{k+1}|\mathsf{L}'_{[2:j]},\mathsf{M}^{(k)})\\
&\overset{(a)}{\geq} \frac{p}{d-k}H(\mathsf{L}_{k+1}|\mathsf{L}_1^{(k)},\mathsf{L}'_{[2:j]},\mathsf{M}^{(k)})\\
&\overset{(b)}{=} \frac{p}{d-k}H(\mathsf{L}_{k+1}|\mathsf{L}_1^{(k)},\mathsf{M}^{(k)})\\
&= \frac{p}{d-k}\left[H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) - H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)})\right], \qquad \text{(A.20)}
\end{aligned}$$

where (a) follows from the fact that conditioning reduces entropy, and (b) is because $\mathsf{L}'_{[2:j]}$ is a function of $\mathsf{L}_1^{(k)}$. Substituting (A.20) into (A.17) gives:

$$\begin{aligned}
&\left(i+\frac{p}{d-k}\right)H(\mathsf{L}_1^{(k)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j)}|\mathsf{M}^{(k)})\\
&\geq \left(i+\frac{p}{d-k}\right)H(\mathsf{L}_1^{(k+1)}|\mathsf{M}^{(k)}) + H(\mathsf{L}_0^{(j-1)}|\mathsf{M}^{(k)}),
\end{aligned}$$

which is equivalent to (A.3) by noting that

$$i+\frac{p}{d-k} = \frac{i(d-k)+p}{d-k} = \frac{d+1-j}{d-k}.$$

This completes the proof of Lemma 12.

APPENDIX B

PROOF OF THE LEMMAS FOR SECURE REGENERATING CODE

B.1 Proof of Lemma 1

Proof. Proof of Lemma 1. Fix $s \in [1:n]$ and $t \in [0:s-1]$. Let us first note that $\underline{\mathsf{S}}_{\to t+1}$ is a function of $\mathsf{W}_{[1:t]}$. As a result, $\mathsf{S}_{\to t+1} = (\underline{\mathsf{S}}_{\to t+1}, \overline{\mathsf{S}}_{\to t+1})$ is a function of $\mathsf{L}_{t,s}$. It thus follows immediately from the node-regeneration requirement (3.3) that W_{t+1} is a function of $\mathsf{L}_{t,s}$. Similarly and inductively, it can be shown that $(\underline{\mathsf{S}}_{\to j}, \mathsf{W}_j)$ is a function of $\mathsf{L}_{t,s}$ for all $j \in [t+2:s]$ as well. This completes the proof of Lemma 1. □

B.2 Proof of Lemma 2

Proof. To prove (3.31), let us fix $t \in [1:2]$, $r \in [2:k-1]$, $p \in [1:r-t+1]$, and $q \in [0:d-r-1]$. We have

$$\begin{aligned}
&H(\mathsf{S}_{1\to[2:p+1]}) + H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+1]\to r+1}) \\
&\overset{(a)}{=} H(\mathsf{S}_{r+q+2\to[r-p+2:r+1]}) + H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+1]\to r+1}) \\
&\overset{(b)}{\geq} H(\mathsf{S}_{r+q+2\to[r-p+2:r]}) + H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+2]\to r+1}) \\
&\overset{(c)}{=} H(\mathsf{S}_{1\to[2:p]}) + H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+2]\to r+1})
\end{aligned}$$

where (a) follows from the fact that $H(\mathsf{S}_{1\to[2:p+1]}) = H(\mathsf{S}_{r+q+2\to[r-p+2:r+1]})$ due to the symmetrical codes that we consider; (b) follows from the submodularity of the entropy function; and (c) follows from the fact that $H(\mathsf{S}_{r+q+2\to[r-p+2:r]}) = H(\mathsf{S}_{1\to[2:p]})$ again due to the symmetrical codes that we consider. This completes the proof of (3.31) for any $r \in [2:k-1]$, $p \in [1:r-t+1]$, and $q \in [0:d-r-1]$.

To prove (3.32), let us fix $t \in [1:2]$, $j \in [2:k]$, and $m \in [1:j-t+1]$. Notice

that (3.32) holds trivially with equality when $j = k$, so we only need to consider the cases where $j \in [2 : k-1]$ for $k \geq 3$. (When $k = 2$, $[2 : k-1]$ is empty and there is nothing to prove.) Now adding (3.31) for $q \in [0 : d-r-1]$ gives:

$$\begin{aligned}
&(d-r)H(\mathsf{S}_{1\to[2:p+1]}) + \sum_{q=0}^{d-r-1} H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+1]\to r+1}) \\
&\geq (d-r)H(\mathsf{S}_{1\to[2:p]}) + \sum_{q=0}^{d-r-1} H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+2]\to r+1}).
\end{aligned}$$

Canceling $\sum_{q=1}^{d-r-1} H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:r+q+1]\to r+1})$ from both sides of the above inequality gives:

$$\begin{aligned}
&(d-r)H(\mathsf{S}_{1\to[2:p+1]}) + H(\mathsf{L}_{t,r}) \\
&\quad \geq (d-r)H(\mathsf{S}_{1\to[2:p]}) + H(\mathsf{L}_{t,r}, \mathsf{S}_{[r+2:n]\to r+1}) \\
&\quad = (d-r)H(\mathsf{S}_{1\to[2:p]}) + H(\mathsf{L}_{t,r+1})
\end{aligned} \tag{B.1}$$

for any $r \in [2 : k-1]$ and $p \in [1 : r-t+1]$. Adding (B.1) for $r \in [j : k-1]$ and $p \in [1 : m]$ gives:

$$\begin{aligned}
&T_{k,d,j} \sum_{p=1}^{m} H(\mathsf{S}_{1\to[2:p+1]}) + m \sum_{r=j}^{k-1} H(\mathsf{L}_{t,r}) \\
&\quad \geq T_{k,d,j} \sum_{p=1}^{m} H(\mathsf{S}_{1\to[2:p]}) + m \sum_{r=j}^{k-1} H(\mathsf{L}_{t,r+1}).
\end{aligned}$$

Canceling $T_{k,d,j} \sum_{p=2}^{m} H(\mathsf{S}_{1\to[2:p]}) + m \sum_{r=j+1}^{k-1} H(\mathsf{L}_{t,r})$ from both sides of the above inequality gives:

$$T_{k,d,j} H(\mathsf{S}_{1\to[2:m+1]}) + mH(\mathsf{L}_{t,j}) \geq mH(\mathsf{L}_{t,k}).$$

Dividing both sides by m completes the proof of (3.32) for any $t \in [1 : 2]$, $j \in [2 : k]$, and

$m \in [1 : j - t + 1]$. This completes the proof of Lemma **??**. □

B.3 Proof of Lemma 3

Proof. To prove (3.33), let us fix $j \in [2 : k-1]$ and $t \in [j : k-1]$. Let $n-j = u(d-t)+p$ for some positive integers u and $p \in [1 : d - t]$. Let

$$\tau_0 := \{1\} \cup [j + 1 : j + p - 1] \tag{B.2}$$

and

$$\tau_q := [j + p + (q - 1)(d - t) : j + p + q(d - t) - 1] \tag{B.3}$$

for $q \in [1 : u - 1]$. Notice that we have

$$\cup_{q=0}^{u-1} \tau_q = \{1\} \cup [j + 1 : t]. \tag{B.4}$$

By the symmetry of the codes that we consider, we have

$$\begin{aligned} &H(\mathsf{S}_{\tau_0 \to t+1} | \mathsf{W}_2, \mathsf{S}_{\to [3:j]}, \mathsf{S}_{t+1,2}) \\ &\qquad = H(\mathsf{S}_{B \to t+1} | \mathsf{W}_2, \mathsf{S}_{\to [3:j]}, \mathsf{S}_{t+1,2}) \end{aligned} \tag{B.5}$$

for any $B \subseteq [t+2:n]$ such that $|B| = |\tau_0| = p$. It follows that

$$\begin{aligned}
&H(\mathsf{S}_{\tau_0 \to t+1} | \mathsf{W}_2, \mathsf{S}_{\to [3:j]}, \mathsf{S}_{t+1 \to 2}) \\
&\overset{(a)}{\geq} \frac{p}{d-t} H(\mathsf{S}_{[t+2:n] \to t+1} | \mathsf{W}_2, \mathsf{S}_{\to [3:j]}, \mathsf{S}_{t+1 \to 2}) \\
&= \frac{p}{d-t} H(\mathsf{S}_{[t+2:n] \to t+1} | \mathsf{W}_2, \underline{\mathsf{S}}_{\to [3:j]}, \overline{\mathsf{S}}_{\to [3:j]}, \mathsf{S}_{t+1 \to 2}) \\
&\overset{(b)}{\geq} \frac{p}{d-t} H(\mathsf{S}_{[t+2:n] \to t+1} | \mathsf{W}_2, \underline{\mathsf{S}}_{\to [3:j]}, \mathsf{L}_{1,t}) \\
&\overset{(c)}{=} \frac{p}{d-t} H(\mathsf{S}_{[t+2:n] \to t+1} | \mathsf{L}_{1,t}) \\
&= \frac{p}{d-t} \left(H(\mathsf{S}_{[t+2:n] \to t+1}, \mathsf{L}_{1,t}) - H(\mathsf{L}_{1,t}) \right) \\
&= \frac{p}{d-t} \left(H(\mathsf{L}_{1,t+1}) - H(\mathsf{L}_{1,t}) \right)
\end{aligned} \tag{B.6}$$

where (a) follows from the well-known Han's inequality [23]; (b) follows from the facts that $(\overline{\mathsf{S}}_{\to [3:j]}, \mathsf{S}_{t+1 \to 2})$ is a sub-collection of random variables from $\mathsf{L}_{1,t}$ and that conditioning reduces entropy; and (c) follows from the fact that $(\mathsf{W}_2, \underline{\mathsf{S}}_{\to [3:j]})$ is a function of $\mathsf{L}_{1,t}$ by Lemma 1.

Next, let us show, by induction, that

$$\begin{aligned}
&rH(\mathsf{L}_{1,t}) + H(\mathsf{L}_{1,j}, \mathsf{S}_{j \to 1}) \\
&\geq rH(\mathsf{L}_{1,t+1}) + H(\mathsf{W}_2, \mathsf{S}_{\to [3:j]}, \mathsf{S}_{\cup_{q=0}^{u-r} \tau_q \to t+1}, \mathsf{S}_{t+1 \to 2})
\end{aligned} \tag{B.7}$$

for any $r \in [1:u]$.

For the base case where $r = 1$, we have

$$
\begin{aligned}
& H(\mathsf{L}_{1,t}) + H(\mathsf{L}_{1,j}, \mathsf{S}_{j\to 1}) \\
& \overset{(a)}{=} H(\mathsf{L}_{1,t}, \mathsf{W}_2, \underline{\mathsf{S}}_{\to[2:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}) + H(\mathsf{L}_{1,j}, \mathsf{S}_{j\to 1}) \\
& \overset{(b)}{=} H(\mathsf{L}_{1,t}, \mathsf{W}_2, \underline{\mathsf{S}}_{\to[2:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}) \\
& \quad + H(\mathsf{L}_{1,j}, \underline{\mathsf{S}}_{\to[2:j-1]}, \mathsf{S}_{j\to 1}) \\
& = H(\mathsf{W}_{[1:2]}, \mathsf{S}_{\to[2:j]}, \overline{\mathsf{S}}_{\to[j+1:t]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}) \\
& \quad + H(\mathsf{W}_1, \mathsf{S}_{\to[2:j-1]}, \overline{\mathsf{S}}_{\to j}, \mathsf{S}_{j\to 1}) \\
& \overset{(c)}{=} H(\mathsf{W}_{[1:2]}, \mathsf{S}_{\to[2:j]}, \overline{\mathsf{S}}_{\to[j+1:t]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}) \\
& \quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}, \overline{\mathsf{S}}_{\to t+1}, \mathsf{S}_{t+1\to 2}) \\
& \overset{(d)}{\geq} H(\mathsf{W}_{[1:2]}, \mathsf{S}_{\to[2:j]}, \overline{\mathsf{S}}_{\to[j+1:t+1]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}) \\
& \quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
& = H(\mathsf{L}_{1,t+1}, \mathsf{W}_2, \underline{\mathsf{S}}_{\to[2:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}) \\
& \quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
& \overset{(e)}{=} H(\mathsf{L}_{1,t+1}) + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2})
\end{aligned}
$$

where (a) and (e) follow from the fact that $(\mathsf{W}_2, \underline{\mathsf{S}}_{\to[2:j]}, \mathsf{S}_{\cup_{q=0}^{u-r}\tau_q \to t+1})$ is a function of $\mathsf{L}_{1,t}$ by Lemma 1; (b) follows from the fact that $\underline{\mathsf{S}}_{\to[2:j]}$ is a function of $\mathsf{L}_{1,j}$ by Lemma 1; and (c) follows from the fact that

$$
\begin{aligned}
& H(\mathsf{W}_1, \mathsf{S}_{\to[2:j-1]}, \overline{\mathsf{S}}_{\to j}, \mathsf{S}_{j\to 1}) \\
& = H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-1}\tau_q \to t+1}, \overline{\mathsf{S}}_{\to t+1}, \mathsf{S}_{t+1\to 2})
\end{aligned} \tag{B.8}
$$

due to the symmetrical codes that we consider; and (d) follows from the submodularity of the entropy function. This completes the proof of the base case.

Now assume that (B.7) holds for some $r \in [1 : u-1]$. Similar to the base case, we have

$$
\begin{aligned}
&H(\mathsf{L}_{1,t}) + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-r}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
&\overset{(a)}{=} H(\mathsf{L}_{1,t}, \mathsf{W}_2, \underline{\mathsf{S}}_{\to[2:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}) \\
&\quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-r}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
&= H(\mathsf{W}_{[1:2]}, \mathsf{S}_{\to[2:j]}, \overline{\mathsf{S}}_{\to[j+1:t]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}) \\
&\quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-r}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
&\overset{(b)}{=} H(\mathsf{W}_{[1:2]}, \mathsf{S}_{\to[2:j]}, \overline{\mathsf{S}}_{\to[j+1:t]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}) \\
&\quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}, \overline{\mathsf{S}}_{\to t+1}, \mathsf{S}_{t+1\to 2}) \\
&\overset{(c)}{\geq} H(\mathsf{W}_{[1:2]}, \mathsf{S}_{\to[2:j]}, \overline{\mathsf{S}}_{\to[j+1:t+1]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}) \\
&\quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
&= H(\mathsf{L}_{1,t+1}, \mathsf{W}_2, \underline{\mathsf{S}}_{\to[2:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}) \\
&\quad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
&\overset{(d)}{=} H(\mathsf{L}_{1,t+1}) + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2})
\end{aligned} \tag{B.9}
$$

where (a) and (d) follow from the fact that $(\mathsf{W}_2, \underline{\mathsf{S}}_{\to[2:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1})$ is a function of $\mathsf{L}_{1,t+1}$ by Lemma 1; (b) follows from the fact that

$$
\begin{aligned}
&H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-r}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}) \\
&= H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}, \overline{\mathsf{S}}_{\to t+1}, \mathsf{S}_{t+1\to 2})
\end{aligned} \tag{B.10}
$$

due to the symmetrical codes that we consider; and (c) follows from the submodularity of

the entropy function. Adding (B.7) and (B.9) gives

$$\begin{aligned}
&(r+1)H(\mathsf{L}_{1,t}) + H(\mathsf{L}_{1,j}, \mathsf{S}_{j\to 1}) \\
&\quad \geq (r+1)H(\mathsf{L}_{1,t+1}) \\
&\qquad + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\cup_{q=0}^{u-(r+1)}\tau_q \to t+1}, \mathsf{S}_{t+1\to 2}).
\end{aligned}$$

This completes the proof of the induction step.

Finally, setting $r = u$ in (B.7) gives:

$$\begin{aligned}
&uH(\mathsf{L}_{1,t}) + H(\mathsf{L}_{1,j}, \mathsf{S}_{j\to 1}) \\
&\geq uH(\mathsf{L}_{1,t+1}) + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{\tau_0\to t+1}, \mathsf{S}_{t+1\to 2}) \\
&= uH(\mathsf{L}_{1,t+1}) + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{t+1\to 2}) \\
&\quad + H(\mathsf{S}_{\tau_0\to t+1}|\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{t+1\to 2}).
\end{aligned} \tag{B.11}$$

Substituting (B.6) into (B.11) and using the fact that

$$u + \frac{p}{d-t} = \frac{u(d-t)+p}{d-t} = \frac{n-j}{d-t} \tag{B.12}$$

we have

$$\begin{aligned}
&\frac{n-j}{d-t}H(\mathsf{L}_{1,t}) + H(\mathsf{L}_{1,j}, \mathsf{S}_{j\to 1}) \\
&\geq \frac{n-j}{d-t}H(\mathsf{L}_{1,t+1}) + H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{t+1\to 2}).
\end{aligned} \tag{B.13}$$

Finally, due to the symmetrical codes that we consider, we have

$$\begin{aligned}
&H(\mathsf{W}_2, \mathsf{S}_{\to[3:j]}, \mathsf{S}_{t+1\to 2}) \\
&\quad = H(\mathsf{W}_1, \mathsf{S}_{\to[2:j-1]}, \mathsf{S}_{j\to 1}) \\
&\quad = H(\mathsf{W}_1, \overline{\mathsf{S}}_{\to[2:j-1]}, \underline{\mathsf{S}}_{\to[2:j-1]}, \mathsf{S}_{j\to 1}) \\
&\quad = H(\mathsf{L}_{1,j-1}, \underline{\mathsf{S}}_{\to[2:j-1]}, \mathsf{S}_{j\to 1}) \\
&\quad = H(\mathsf{L}_{1,j-1}, \mathsf{S}_{j\to 1}) \qquad \text{(B.14)}
\end{aligned}$$

where the last equality follows from the fact that $\underline{\mathsf{S}}_{\to[2:j-1]}$ is a function of $\mathsf{L}_{1,j-1}$ by Lemma 1. Substituting (B.14) into (B.13) completes the proof of (3.33) for any $j \in [2 : k-1]$ and $t \in [j : k-1]$. □

B.4 Proof of Lemma 4

Proof. First note that

$$\begin{aligned}
&\sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) \\
&\overset{(a)}{=} \sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}, \mathsf{W}_i|\mathsf{W}_{[1:i-1]}) \\
&= \sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i]}) + \sum_{i=1}^{n-1} H(\mathsf{W}_i|\mathsf{W}_{[1:i-1]}) \\
&= \sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i]}) + H(\mathsf{W}_{[1:n-1]}) \\
&= \sum_{i=1}^{n-1} H(\mathsf{S}_{[i+1:n]\to i}|\mathsf{W}_{[1:i]}) + H(\mathsf{W}_{[1:n-1]}) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} H(\mathsf{S}_{j\to i}|\mathsf{W}_{[1:i]}, \mathsf{S}_{[i+1;j-1]\to i}) + H(\mathsf{W}_{[1:n-1]})
\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\geq} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} H(\mathsf{S}_{j\to i}|\mathsf{W}_{[1:j-1]}) + H(\mathsf{W}_{[1:n-1]}) \\
&= \sum_{j=2}^{n} \sum_{i=1}^{j-1} H(\mathsf{S}_{j\to i}|\mathsf{W}_{[1:j-1]}) + H(\mathsf{W}_{[1:n-1]}) \\
&\geq \sum_{j=2}^{n-1} \sum_{i=1}^{j-1} H(\mathsf{S}_{j\to i}|\mathsf{W}_{[1:j-1]}) + H(\mathsf{W}_{[1:n-1]}) \\
&\geq \sum_{j=2}^{n-1} H(\mathsf{S}_{j\to [1:j-1]}|\mathsf{W}_{[1:j-1]}) + H(\mathsf{W}_{[1:n-1]}) \\
&= \sum_{j=2}^{n-1} H(\mathsf{S}_{j\to [1:j-1]}) - \sum_{j=1}^{n-1} I(\mathsf{S}_{j\to [1:j-1]}; \mathsf{W}_{[1:j-1]}) \\
&\quad + H(\mathsf{W}_{[1:n-1]}) \\
&\overset{(c)}{=} \sum_{j=2}^{n-1} H(\mathsf{S}_{j+1\to [2:j]}) - \sum_{j=1}^{n-1} I(\mathsf{S}_{j\to [1:j-1]}; \mathsf{W}_{[1:j-1]}) \\
&\quad + H(\mathsf{W}_{[1:n-1]}) \\
&\geq H(\overline{\mathsf{S}}_{\to [2:n-1]}) - \sum_{j=1}^{n-1} I(\mathsf{S}_{j\to [1:j-1]}; \mathsf{W}_{[1:j-1]}) \\
&\quad + H(\mathsf{W}_{[1:n-1]})
\end{aligned} \tag{B.15}
$$

where (a) follows from the fact that W_i is a function of $(\mathsf{W}_{[1:i-1]}, \overline{\mathsf{S}}_{\to i}) = \mathsf{L}_{i-1,i}$ by Lemma 1; (b) follows from the fact that $\mathsf{S}_{[i+1;j-1]\to i}$ is a function of $\mathsf{W}_{[1:j-1]}$; and (c) follows from the fact that $H(\mathsf{S}_{j\to [1:j-1]}) = H(\mathsf{S}_{j+1\to [2:j]})$ due to the symmetrical code that we consider.

Further note that

$$
\begin{aligned}
&H(\overline{\mathsf{S}}_{\to [2:n-1]}) + H(\mathsf{W}_1) \\
&\qquad \geq H(\mathsf{W}_1, \overline{\mathsf{S}}_{\to [2:n-1]}) = H(\mathsf{L}_{1,n-1}).
\end{aligned} \tag{B.16}
$$

Adding (B.15)–(B.16) gives:

$$\begin{aligned}
&\sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + H(\mathsf{W}_1) \\
&\geq H(\mathsf{L}_{1,n-1}) + H(\mathsf{W}_{[1:n-1]}) - \sum_{j=1}^{n-1} I(\mathsf{S}_{j\to[1:j-1]}; \mathsf{W}_{[1:j-1]}) \\
&\overset{(a)}{\geq} H(\mathsf{L}_{1,n-1}) + H(\mathsf{W}_{[1:n-1]}) - \sum_{j=1}^{n-1} I(\mathsf{W}_j; \mathsf{W}_{[1:j-1]}) \\
&= H(\mathsf{L}_{1,n-1}) + H(\mathsf{W}_{[1:n-1]}) \\
&\quad - \sum_{j=1}^{n-1} \big(H(\mathsf{W}_j) + H(\mathsf{W}_{[1:j-1]}) - H(\mathsf{W}_{[1:j]})\big) \\
&= H(\mathsf{L}_{1,n-1}) + 2H(\mathsf{W}_{[1:n-1]}) - \sum_{j=1}^{n-1} H(\mathsf{W}_j) \\
&\overset{(b)}{\geq} 3H(\mathsf{W}_{[1:n-1]}) - \sum_{j=1}^{n-1} H(\mathsf{W}_j)
\end{aligned} \tag{B.17}$$

where (a) follows from the fact that $\mathsf{S}_{j\to[1:j-1]}$ is a function of W_j; and (b) follows from the fact that $\mathsf{W}_{[2:n-1]}$ is a function of $\mathsf{L}_{1,n-1}$ by Lemma 1 so we have $H(\mathsf{L}_{1,n-1}) \geq H(\mathsf{W}_{[1:n-1]})$.

Finally, by the node-capacity constraint, we have

$$n\alpha \geq H(\mathsf{W}_1) + \sum_{j=1}^{n-1} H(\mathsf{W}_j). \tag{B.18}$$

Adding (B.17) and (B.18) gives:

$$\begin{aligned}
&\sum_{i=1}^{n-1} H(\overline{\mathsf{S}}_{\to i}|\mathsf{W}_{[1:i-1]}) + n\alpha \\
&\geq 3H(\mathsf{W}_{[1:n-1]}) \overset{(a)}{=} 3H(\mathsf{W}_{[1:n]}, \mathsf{M}) \overset{(b)}{=} 3H(\mathsf{W}_{[1:n]}, \mathsf{M}, \mathsf{S}_{\to 1}) \\
&\geq 3H(\mathsf{M}, \mathsf{S}_{\to 1}) = 3H(\mathsf{S}_{\to 1}) + 3H(\mathsf{M}|\mathsf{S}_{\to 1}) \\
&\overset{(c)}{=} 3H(\mathsf{S}_{\to 1}) + 3H(\mathsf{M}) = 3H(\mathsf{S}_{\to 1}) + 3B
\end{aligned}$$

where (a) follows from the facts that M is a function of $\mathsf{W}_{[1:n-1]}$ by the message-recovery requirement (3.2) and that W_n is a function of $\mathsf{S}_{\to n}$, which is in turn a function of $\mathsf{W}_{[1:n-1]}$; (b) follows from the fact that $\mathsf{S}_{\to 1}$ is a function of $\mathsf{W}_{[2:n]}$; and (c) follows from the repair-secrecy requirement (3.4) with $\ell = 1$. This completes the proof of Lemma 4. $\square$

APPENDIX C

PROOF OF THE LEMMAS FOR CACHING PROBLEM

C.1 Proof of Lemma 6

Proof. In this proof, we need to use induction. Obviously (4.27) holds for $p = 1$. We assume when it holds for $p = k$, then, and for $p = k + 1 \leq N - 1$, we have

$$
\begin{aligned}
& H(Z_1) + (k+1)H(X_{2,K}) \\
\geq & H(Z_1, W_{[k]}) + kH(X_{k+1,K}|Z_1, W_{[k]}) + H(X_{2,K}) && \text{(C.1)} \\
= & H(Z_1, W_{[k]}) + kH(X_{k+2,K}|Z_1, W_{[k]}) + H(X_{k+1,k+2,K}) && \text{(C.2)} \\
\geq & H(Z_1, W_{[k+1]}) + H(X_{k+1,k+2,K}|Z_1, W_{[k+1]}) + kH(X_{k+2,K}|Z_1, W_{[k]}) && \text{(C.3)} \\
= & H(Z_1, W_{[k+1]}) + H(X_{k+2,K}|Z_1, W_{[k+1]}) + kH(X_{k+2,K}|Z_1, W_{[k]}) && \text{(C.4)} \\
\geq & H(Z_1, W_{[k+1]}) + (k+1)H(X_{k+2,K}|Z_1, W_{[k]}). && \text{(C.5)}
\end{aligned}
$$

In $(C.2)$, $H(X_{k+1,K}|Z_1, W_{[k]}) = H(X_{k+2,K}|Z_1, W_{[k]})$ is because of the symmetry under file index permutation $k + 1 \leftrightarrow k + 2$. Similarly, we have $H(X_{2,K}) = H(X_{k+1,k+2,K})$ because of file index permutation $1 \leftrightarrow k + 1, 2 \leftrightarrow k + 2$.

In $(C.4)$, $H(X_{k+1,k+2,K}|Z_1, W_{[k+1]}) = H(X_{k+2,K}|Z_1, W_{[k+1]})$ is also because of file index permutation: switching file index 1 and $k + 1$ can help us get this.

This induction gives us that inequality (4.27) holds for every $p \leq N-1$. When $p = N$,

inequality (4.27) is degraded as

$$\begin{aligned} & H(Z_1) + NH(X_{2,K}) \\ =& H(Z_K) + H(X_{[2:N],K}) + H(X_{2,1,K}) && \text{(C.6)} \\ \geq& H(Z_K, W_{[N]}) && \text{(C.7)} \\ =& H(Z_1, W_{[N]}) && \text{(C.8)} \end{aligned}$$

In $(C.6)$, we have

$$H(X_{2,K}) = H(X_{3,K}) = \cdots = H(X_{N,K})$$

according to file index symmetry and so is the reason for $H(X_{2,K}) = H(X_{2,1,K})$. In $(C.8)$, $H(Z_K, W_{[N]}) = H(Z_1, W_{[N]})$ is because of the symmetry for user index, a permutation switch user index 1 and K will help us get that.

Combine them both we have the whole proof for Lemma 6. □

C.2 Proof of Lemma 7

Proof. When $K = 2$, this inequality trivially degraded to

$$H(W_{[p]}) \geq pF, \tag{C.9}$$

which is obvious. Therefore we only need to focus on $K \geq 3$.

Firstly, we can see

$$
\begin{aligned}
&(K-2)H(Z_1, W_{[p]}) \\
&+ \frac{(K-2)(K-3)}{2}(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \\
\geq &(K-2)H(Z_1, W_{[p]}) \\
&+ \frac{(K-2)(K-3)}{2}H(X_{[p+1:N],K}|Z_1, W_{[p]}) && \text{(C.10)} \\
= &\sum_{i=1}^{K-2}[H(Z_1, W_{[p]}) + (i-1)H(X_{[p+1:N],K}|Z_1, W_{[p]})] && \text{(C.11)} \\
= &\sum_{i=1}^{K-2}[H(Z_1, W_{[p]}) + (i-1)H(X_{[p+1:N],K}|Z_1, W_{[p]})] && \text{(C.12)} \\
\geq &\sum_{i=1}^{K-2}[H(Z_{K-i+1}, W_{[p]}) + H(X_{[p+1:N],[K-i+2:K]}|Z_{K-i+1}, W_{[p]})] && \text{(C.13)} \\
= &\sum_{i=1}^{K-2} H(Z_{K-i+1}, W_{[p]}, X_{[p+1:N],[K-i+2:K]}). && \text{(C.14)}
\end{aligned}
$$

Need to mention that when $i = 1$, we regard $X_{[p+1:N],[K-i+2:K]}$ as an empty set. Also, in formula $(C.13)$, we have $(i-1)H(X_{[p+1:N],K}|Z_1, W_{[p]}) \geq H(X_{[p+1:N],[K-i+2:K]}|Z_{K-i+1}, W_{[p]})$ because for each $k \in [K-i+2:K]$, as long as $[K-i+2:K]$ is not empty, we have $H(X_{[p+1:N],K}|Z_1, W_{[p]}) = H(X_{[p+1:N],k}|Z_1, W_{[p]})$ according to the symmetry of user index.

Consider following inequality:

$$
\begin{aligned}
&H(W_{[p]}, X_{[p+1:N],[k:K]}) + H(Z_k, W_{[p]}, X_{[p+1:N],[k+1:K]}) \\
\geq &H(Z_k, W_{[p]}, X_{[p+1:N],[k:K]}) + H(W_{[p]}, X_{[p+1:N],[k+1:K]}) \\
\geq &NF + H(W_{[p]}, X_{[p+1:N],[k+1:K]}) && \text{(C.15)}
\end{aligned}
$$

If we sum up all $k \in [3:K-1]$, we have

$$\begin{aligned}
&\sum_{k=3}^{K-1}\left[H(W_{[p]}, X_{[p+1:N],[k:K]}) + H(Z_k, W_{[p]}, X_{[p+1:N],[k+1:K]})\right] \\
\geq &\sum_{k=3}^{K-1}\left(NF + H(W_{[p]}, X_{[p+1:N],[k+1:K]})\right) \qquad \text{(C.16)}
\end{aligned}$$

Notice that

$$\sum_{k=4}^{K-1} H(W_{[p]}, X_{[p+1:N],[k:K]}) = \sum_{k=3}^{K-2} H(W_{[p]}, X_{[p+1:N],[k+1:K]}), \qquad \text{(C.17)}$$

which means that they can be canceled from both sides of inequality. Meanwhile, we can rewrite our expression as

$$\begin{aligned}
&\sum_{k=3}^{K-1} H(Z_k, W_{[p]}, X_{[p+1:N],[k+1:K]}) \\
= &\sum_{i=2}^{K-2} H(Z_{K-i+1}, W_{[p]}, X_{[p+1:N],[K-i+2:K]}). \qquad \text{(C.18)}
\end{aligned}$$

Take $(C.18)$ back to $(C.16)$, we have

$$\begin{aligned}
&H(W_{[p]}, X_{[p+1:N],[3:K]}) + \sum_{i=2}^{K-2} H(Z_{K-i+1}, W_{[p]}, X_{[p+1:N],[K-i+2:K]}) \\
\geq &(K-3)NF + H(W_{[p]}, X_{[p+1:N],K}). \qquad \text{(C.19)}
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&H(W_{[p]}, X_{[p+1:N],K}) + H(Z_K, W_{[p]}) \\
\geq &H(Z_K, W_{[p]}, X_{[p+1:N],K}) + H(W_{[p]}) \\
\geq &NF + pF \qquad \text{(C.20)}
\end{aligned}$$

Combine $(C.19)$ and $(C.19)$, we have

$$
\begin{aligned}
&H(W_{[p]}, X_{[p+1:N],[3:K]}) + (K-2)H(Z_1, W_{[p]}) \\
&+ \frac{(K-2)(K-3)}{2}(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \\
\geq &H(W_{[p]}, X_{[p+1:N],[3:K]}) + \sum_{i=1}^{K-2} H(Z_{K-i+1}, W_{[p]}, X_{[p+1:N],[K-i+2:K]}) \qquad \text{(C.21)} \\
= &H(W_{[p]}, X_{[p+1:N],[3:K]}) + \sum_{i=2}^{K-2} H(Z_{K-i+1}, W_{[p]}, X_{[p+1:N],[K-i+2:K]}) + H(Z_K, W_{[p]}) \\
& \qquad \text{(C.22)} \\
\geq &(K-3)NF + H(W_{[p]}, X_{[p+1:N],K}) + H(Z_K, W_{[p]}) \qquad \text{(C.23)} \\
\geq &(K-3)NF + NF + pF \qquad \text{(C.24)} \\
= &(K-2)NF + pF \qquad \text{(C.25)}
\end{aligned}
$$

Hence Lemma 2 is proved. □

C.3 Proof of Lemma 8

Proof. When $K = 2$, the proof is trivial as following:

When $K = 2$, $X_{[p+1:N],[3:2]}$ and $X_{[p+1:N],[3:2]}$ are empty sets. Hence Lemma 8 is degrade as

$$
\begin{aligned}
&2H(Z_1, W_{[p]}) + (N-p)H(X_{p+1,2}|Z_1, W_{[p]}) \\
\geq &H(Z_{[2]}, W_{[p]}) + (N-p)H(X_{p+1,2}|Z_1, W_{[p]}) + H(W_{[p]}) \\
\geq &H(Z_{[2]}, W_{[p]}, X_{[p+1:N],2}) + H(W_{[p]}) \\
\geq &N + p \qquad \text{(C.26)}
\end{aligned}
$$

Therefore, now we put our focus on the case that $K \geq 3$. According to symmetry in file index, we have following claims:

- For any $p+1 \leq N$, we have:

$$(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \geq H(X_{[p+1:N],K}|Z_1, W_{[p]}) \tag{C.27}$$

where (C.27) follows the fact that

$$H(X_{p+1,K}|Z_1, W_{[p]}) = H(X_{i,K}|Z_1, W_{[p]}) \tag{C.28}$$

for any $i \in [p+1:N]$.

- for any $i, j \in [k-1]$ and $p+1 \leq N$, we have

$$H(X_{[p+1:N],k}|Z_i, W_{[p]}) = H(X_{[p+1:N],k}|Z_j, W_{[p]}) \tag{C.29}$$

where (C.29) follows a permutation on the user index that $\pi : i \leftrightarrow j$.

- For any $k_1, k_2 \in [i+1:K]$ and $p+1 \leq N$, we have

$$H(X_{[p+1:N],k_1}|Z_i, W_{[p]}) = H(X_{[p+1:N],k_2}|Z_i, W_{[p]}) \tag{C.30}$$

where (C.30) follows a permutation on the user index that $\pi : k_1 \leftrightarrow k_2$.

With all claims above, we have

$$
\begin{aligned}
&H(Z_1, W_{[p]}) + (K-2)(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \\
\overset{(a)}{\geq} &H(Z_1, W_{[p]}) + (K-2)H(X_{[p+1:N],K}|Z_1, W_{[p]}) \\
\overset{(b)}{=} &H(Z_1, W_{[p]}) + \sum_{k=3}^{K} H(X_{[p+1:N],k}|Z_1, W_{[p]}) \\
\geq &H(Z_1, W_{[p]}) + H(X_{[p+1:N],[3:K]}|Z_1, W_{[p]}) \\
= &H(Z_1, W_{[p]}, X_{[p+1:N],[3:K]}) \qquad \text{(C.31)}
\end{aligned}
$$

(a) follows (C.27) and (b) follows (C.29).

Further more,

$$
\begin{aligned}
&2H(Z_1, W_{[p]}, X_{[p+1:N],[3:K]}) \\
= &H(Z_1, W_{[p]}, X_{[p+1:N],[3:K]}) + H(Z_2, W_{[p]}, X_{[p+1:N],[3:K]}) \qquad \text{(C.32)} \\
\geq &H(Z_{[2]}, W_{[p]}, X_{[p+1:N],[3:K]}) + H(W_{[p]}, X_{[p+1:N],[3:K]}) \qquad \text{(C.33)} \\
\geq &H(Z_{[2]}, W_{[p]}, X_{[p+1:N],3}) + H(W_{[p]}, X_{[p+1:N],[3:K]}) \qquad \text{(C.34)}
\end{aligned}
$$

$H(Z_{[2]}, W_{[p]}, X_{[p+1:N],3})$, the first item in (C.34), satisfies

$$
\begin{aligned}
&H(Z_{[2]}, W_{[p]}, X_{[p+1:N],3}) + (N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \\
= &H(Z_{[2]}, W_{[p]}, X_{[p+1:N],3}) + (N-p)H(X_{p+1,2}|Z_1, W_{[p]}) \qquad \text{(C.35)} \\
\geq &H(Z_{[2]}, W_{[p]}, X_{[p+1:N],3}) + H(X_{[p+1:N],2}|Z_1, W_{[p]}) \qquad \text{(C.36)} \\
\geq &H(Z_{[2]}, W_{[p]}, X_{[p+1:N],2}, X_{[p+1:N],3}) \qquad \text{(C.37)} \\
\geq &N, \qquad \text{(C.38)}
\end{aligned}
$$

and $H(W_{[p]}, X_{[p+1:N],[3:K]})$, the second item in (C.34), satisfies

$$\begin{aligned} & H(W_{[p]}, X_{[p+1:N],[3:K]}) + (K-2)H(Z_1, W_{[p]}) \\ & + \frac{(K-2)(K-3)}{2}(N-p)H(X_{p+1,K}|Z_1, W_{[p]}) \\ \geq & (K-2)N + p \end{aligned} \tag{C.39}$$

(C.39) follows Lemma (7). By adding (C.31)), (C.38)) and (C.39)) together, we have Lemma 8. □

C.4 Proof of Lemma 9

Proof. When A=1, we have

$$\begin{aligned} & H(X_{(i+1,k+1)}|Z_{[k]}, W_{[i]}) + H(Z_{[k+1]}, W_{[i]}) \\ = & H(Z_{[k+1]}, W_{[i]}, X_{(i+1,k+1)}) && \text{(C.40)} \\ = & H(Z_{[k+1]}, W_{[i+1]}, X_{(i+1,k+1)}) && \text{(C.41)} \\ = & H(X_{(i+1,k+1)}|Z_{[k+1]}, W_{[i+1]}) + H(Z_{[k+1]}, W_{[i+1]}) && \text{(C.42)} \\ = & H(X_{(i+2,k+2)}|Z_{[k+1]}, W_{[i+1]}) + H(Z_{[k+1]}, W_{[i+1]}) && \text{(C.43)} \end{aligned}$$

In $(C.43)$, we have $H(X_{(i+1,k+1)}|Z_{[k+1]}, W_{[i+1]}) = H(X_{(i+2,k+2)}|Z_{[k+1]}, W_{[i+1]})$ by switching file index $i+2$ and $N-K+(k+2)$, $i+1$ and $k+1$. More precisely, we have

$$X_{(i+1,k+1)} = X_{(\underbrace{1, 2, \ldots, k}_{k \quad users}, i+1, \underbrace{N-K+(k+2), \ldots, N}_{K-k-1 \quad users})} \tag{C.44}$$

$$X_{(i+2,k+2)} = X_{(\underbrace{1, 2, \ldots, k, k+1}_{k+1 \quad users}, i+2, \underbrace{N-K+(k+3), \ldots, N}_{K-k-2 \quad users})}. \tag{C.45}$$

By comparison, we can find see the corresponding permutation clearly. Now, we assume (4.51) holds for $A = m$. When $A = m + 1 \leq N - (K - k) - 1$, we have

$$
\begin{aligned}
&(m+1)H(X_{(i+1,k+1)}|Z_{[k]}, W_{[i]}) + H(Z_{[k+1]}, W_{[i]}) \\
\geq &mH(X_{(i+m+1,k+2)}|Z_{[k+1]}, W_{[i+m]}) + H(Z_{[k+1]}, W_{[i+m]}) + H(X_{(i+1,k+1)}|Z_{[k]}, W_{[i]}) \qquad \text{(C.46)} \\
= &mH(X_{(i+m+1,k+2)}|Z_{[k+1]}, W_{[i+m]}) + H(Z_{[k+1]}, W_{[i+m]}) + H(X_{(i+m+1,k+1)}|Z_{[k]}, W_{[i]}) \qquad \text{(C.47)} \\
\geq &mH(X_{(i+m+1,k+2)}|Z_{[k+1]}, W_{[i+m]}) + H(Z_{[k+1]}, W_{[i+m+1]}) \\
&+ H(X_{(i+m+1,k+1)}|Z_{[k+1]}, W_{[i+m+1]}) \qquad \text{(C.48)} \\
= &mH(X_{(i+m+1,k+2)}|Z_{[k+1]}, W_{[i+m]}) + H(Z_{[k+1]}, W_{[i+m+1]}) \\
&+ H(X_{(i+m+2,k+2)}|Z_{[k+1]}, W_{[i+m+1]}) \qquad \text{(C.49)} \\
\geq &mH(X_{(i+m+2,k+2)}|Z_{[k+1]}, W_{[i+m+1]}) + H(Z_{[k+1]}, W_{[i+m+1]}) \\
&+ H(X_{(i+m+2,k+2)}|Z_{[k+1]}, W_{[i+m+1]}) \qquad \text{(C.50)} \\
= &(m+1)H(X_{(i+m+2,k+2)}|Z_{[k+1]}, W_{[i+m+1]}) + H(Z_{[k+1]}, W_{[i+m+1]}) \qquad \text{(C.51)}
\end{aligned}
$$

In $(C.49)$,

$$H(X_{(i+m+1,k+1)}|Z_{[k+1]}, W_{[i+m+1]}) = H(X_{(i+m+2,k+2)}|Z_{[k+1]}, W_{[i+m+1]})$$

because of switching $i + m + 2$ and $N - K + (k + 2)$.

By induction we proved this lemma. □

C.5 Proof of Lemma 10

Proof. As our definition, we have

$$k \leq p_k \leq k + (N - K) - 1 \tag{C.52}$$

First half is because of

$$\begin{aligned} & p_k \\ =& \lfloor \frac{k(k+1)N}{K(K+1)} \rfloor & \text{(C.53)} \\ \geq& \lfloor \frac{k(k+1)}{2} \rfloor & \text{(C.54)} \\ \geq& k. & \text{(C.55)} \end{aligned}$$

For the second half, We obviously have that

$$p_k = \lfloor \frac{k(k+1)N}{K(K+1)} \rfloor \leq \frac{k(k+1)N}{K(K+1)}. \tag{C.56}$$

Further more, since

$$\begin{aligned} & N - \frac{k(k+1)N}{K(K+1)} - (K - k) \\ =& \frac{K(K+1) - k(k+1)}{K(K+1)} N - (K - k) & \text{(C.57)} \\ =& \frac{(K-k)(K+k+1)}{K(K+1)} N - (K - k) & \text{(C.58)} \\ =& (K-k)(\frac{(K+k+1)}{K(K+1)} N - 1) & \text{(C.59)} \\ \geq& (K-k)(\frac{(K+k+1)}{2} - 1) & \text{(C.60)} \\ \geq& 1, & \text{(C.61)} \end{aligned}$$

that is to say,

$$\frac{k(k+1)N}{K(K+1)} \leq k + (N-K) - 1 \tag{C.62}$$

Hence $(C.52)$ is proved. $(C.52)$ ensures that every $X_{(p_k+1,k+1)}$ and $X_{(p_k+2,k+1)}$ is legitimate by our definition, thus we can make sure $p_k \geq k$ and $p_k + 1 \leq k + 1 + (N-K)$. Now we go back to our proof. We use induction to prove (4.54). Firstly, when $k = 1$, we have

$$\begin{aligned}
&(K+1)!MF + (K-1)!2NRF \\
=&(K-1)![K(K+1) - q_1][MF + p_1RF] \\
&+ (K-1)!q_1[NF + (p_1+1)RF] &\text{(C.63)}\\
\geq&(K-1)![K(K+1) - q_1][H(Z_1) + p_1H(X_{1,2,\ldots,K})] \\
&+ (K-1)!q_1[H(Z_1) + (p_1+1)H(X_{1,2,\ldots,K})] &\text{(C.64)}\\
\geq&(K-1)![K(K+1) - q_1][H(Z_1, W_{[p_1]}) + p_1H(X_{(p_1+1,2)}|Z_1, W_{[p_1]})] \\
&+ (K-1)!q_1[H(Z_1, W_{[p_1+1]}) + p_1H(X_{(p_1+1,2)}|Z_1, W_{[p_1]}) + H(X_{(p_1+2,2)}|Z_1, W_{[p_1+1]})] \\
&&\text{(C.65)}\\
=&(K-1)![(K(K+1) - q_1)H(Z_1, W_{[p_1]}) + q_1H(Z_1, W_{[p_1+1]})] \\
&+ (K-1)![(K(K+1)p_1H(X_{(p_1+1,2)}|Z_1, W_{[p_1]}) + q_1H(X_{(p_1+2,2)}|Z_1, W_{[p_1+1]})] \\
&&\text{(C.66)}
\end{aligned}$$

Need to explain that in $(C.65)$,

$$\begin{aligned}
&H(Z_1) + p_1H(X_{1,2,\ldots,K}) \\
\geq&H(Z_1, W_{[p_1]}) + p_1H(X_{(p_1+1,2)}|Z_1, W_{[p_1]}) &\text{(C.67)}
\end{aligned}$$

is by Lemma 9 with setting $A = p_1$ and $i = 0$, and

$$\begin{aligned}
&H(Z_1) + (p_1 + 1)H(X_{1,2,\ldots,K}) \\
\geq &H(Z_1, W_{[p_1+1]}) + p_1 H(X_{(p_1+1,2)}|Z_1, W_{[p_1]}) + H(X_{(p_1+2,2)}|Z_1, W_{[p_1+1]})
\end{aligned} \tag{C.68}$$

is by applying lemma 1 twice, setting $A = p_1$, $i = 0$ and $A = 1$, $i = p_1$ respectively.

If we assume (4.54) holds for some k, then for $k + 1$, we have

$$\begin{aligned}
&(K+1)!MF + (K-1)!2NRF \\
\geq &\frac{(K-1)!}{k}[(K(K+1) - q_k)H(Z_{[k]}, W_{[p_k]}) + q_k H(Z_{[k]}, W_{[p_k+1]})] \\
&+ \frac{(K-1)!}{k}[(2kN - q_k)H(X_{(p_k+1,k+1)}|Z_{[k]}, W_{[p_k]}) \\
&\quad + q_k H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]})] \\
&+ (k-1)(K-1)!NF \\
\geq &\frac{(K-1)!}{k}\frac{k}{k+1}[(K(K+1) - q_k)H(Z_{[k+1]}, W_{[p_k]}) + q_k H(Z_{[k+1]}, W_{[p_k+1]})] \\
&+ \frac{(K-1)!}{k}\frac{1}{k+1}[(K(K+1) - q_k)H(W_{[p_k]}) + q_k H(W_{[p_k+1]})] \\
&+ \frac{(K-1)!}{k}[(2kN - q_k)H(X_{(p_k+1,k+1)}|Z_{[k]}, W_{[p_k]}) \\
&\quad + q_k H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]})] \\
&+ (k-1)(K-1)!NF
\end{aligned} \tag{C.69}$$

$(C.69)$ is according to Han's Inequality as following:

$$
\begin{aligned}
& H(Z_{[k]}, W_{[p]}) \\
=& H(Z_{[k]}|W_{[p]}) + H(W_{[p]}) \\
\geq& \frac{k}{k+1} H(Z_{[k+1]}|W_{[p]}) + H(W_{[p]}) && \text{(C.70)} \\
=& \frac{k}{k+1} H(Z_{[k+1]}, W_{[p]}) + \frac{1}{k+1} H(W_{[p]}) && \text{(C.71)}
\end{aligned}
$$

Since $H(W_{[p]}) = pF$ as our definition, we can simplify some items in $(C.69)$:

$$
\begin{aligned}
& \frac{(K-1)!}{k} \frac{1}{k+1} [(K(K+1) - q_k) H(W_{[p_k]}) + q_k H(W_{[p_k+1]})] \\
=& \frac{(K-1)!}{k} \frac{1}{k+1} [K(K+1) p_k F + q_k F] \\
=& \frac{(K-1)!}{k} \frac{1}{k+1} [k(k+1) NF] \\
=& (K-1)! NF, && \text{(C.72)}
\end{aligned}
$$

This result can be combined with item $(k-1)(K-1)!NF$. For the remaining part, according to Lemma 9, we have following separated steps:

Firstly,

$$
\begin{aligned}
& \frac{(K-1)!}{k} \frac{k}{k+1} [(K(K+1) - q_k) H(Z_{[k+1]}, W_{[p_k]}) + q_k H(Z_{[k+1]}, W_{[p_k+1]})] \\
& + \frac{(K-1)!}{k+1} (K(K+1) - q_k) H(X_{(p_k+1, k+1)} | Z_{[k]}, W_{[p_k]}) \\
\geq& \frac{(K-1)!}{k+1} K(K+1) H(Z_{[k+1]}, W_{[p_k+1]}) \\
& + \frac{(K-1)!}{k+1} (K(K+1) - q_k) H(X_{(p_k+2, k+2)} | Z_{[k+1]}, W_{[p_k+1]}) && \text{(C.73)} \\
\geq& \frac{(K-1)!}{k+1} K(K+1) H(Z_{[k+1]}, W_{[p_k+1]}) \\
& + \frac{(K-1)!}{k+1} (K(K+1) - q_k) H(X_{(p_{k+1}+1, k+2)} | Z_{[k+1]}, W_{[p_{k+1}]}). && \text{(C.74)}
\end{aligned}
$$

by setting $A = 1$ and $i = p_k$. In $(C.74)$, we have

$$\begin{aligned}
&H(X_{(p_k+2,k+2)}|Z_{[k+1]}, W_{[p_k+1]}) \\
=&H(X_{(p_{k+1}+1,k+2)}|Z_{[k+1]}, W_{[p_k+1]}) \\
\geq&H(X_{(p_{k+1}+1,k+2)}|Z_{[k+1]}, W_{[p_{k+1}]})
\end{aligned}$$

because of symmetry under permutation.

Secondly, we have

$$\begin{aligned}
&\frac{(K-1)!}{k+1}K(K+1)H(Z_{[k+1]}, W_{[p_k+1]}) \\
&+\frac{(K-1)!}{k+1}K(K+1)(p_{k+1}-p_k-1)H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]}) \\
\geq&\frac{(K-1)!}{k+1}K(K+1)H(Z_{[k+1]}, W_{[p_{k+1}]}) \\
&+\frac{(K-1)!}{k+1}K(K+1)(p_{k+1}-p_k-1)H(X_{(p_{k+1}+1,k+2)}|Z_{[k+1]}, W_{[p_{k+1}]})
\end{aligned} \tag{C.75}$$

according to Lemma 9 as well by setting $A = p_{k+1} - p_k - 1$ and $i = p_k + 1$, and at last for the third step, we have

$$\begin{aligned}
&\frac{(K-1)!}{k+1}q_{k+1}H(Z_{[k+1]}, W_{[p_{k+1}]}) \\
&+\frac{(K-1)!}{k+1}q_{k+1}H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]}) \\
\geq&\frac{(K-1)!}{k+1}q_{k+1}H(Z_{[k+1]}, W_{[p_{k+1}+1]}) \\
&+\frac{(K-1)!}{k+1}q_{k+1}H(X_{(p_{k+1}+2,k+2)}|Z_{[k]}, W_{[p_{k+1}+1]})
\end{aligned} \tag{C.76}$$

by setting $A = 1$ and $i = p_{k+1}$.

From $(C.74)$ to $(C.76)$, these three steps turn every$H(Z_{[k+1]}, W_{[p_k]})$ into $H(Z_{[k+1]}, W_{[p_{k+1}]})$ or $H(Z_{[k+1]}, W_{[p_{k+1}+1]})$, and turn all $H(X_{(p_k+1,k+1)}|Z_{[k]}, W_{[p_k]})$ and $H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]})$

into $H(X_{(p_{k+1}+1,k+2)}|Z_{[k+1]}, W_{[p_{k+1}]})$ and $H(X_{(p_{k+1}+2,k+2)}|Z_{[k+1]}, W_{[p_{k+1}+1]})$. We will verify the coefficient of these three steps as following.

The summation of those coefficients is like:

$$
\begin{aligned}
&\frac{(K-1)!}{k+1}(K(K+1)-q_k) \\
+&\frac{(K-1)!}{k+1}K(K+1)(p_{k+1}-p_k-1) \\
+&\frac{(K-1)!}{k+1}q_{k+1} \\
=&\frac{(K-1)!}{k+1}[K(K+1)-q_k+K(K+1)(p_{k+1}-p_k)-K(K+1)+q_{k+1}] && \text{(C.77)} \\
=&\frac{(K-1)!}{k+1}[K(K+1)(p_{k+1}-p_k)+q_{k+1}-q_k] && \text{(C.78)} \\
=&\frac{(K-1)!}{k+1}[((k+1)(k+2)N-q_{k+1})-(k(k+1)N-q_k)+q_{k+1}-q_k] && \text{(C.79)} \\
=&\frac{(K-1)!}{k+1}2(k+1)N && \text{(C.80)} \\
=&(K-1)!2N && \text{(C.81)}
\end{aligned}
$$

Since we have

$$
\begin{aligned}
&kK(K+1)+q_k \\
\leq&kK(K+1)+K(K+1) && \text{(C.82)} \\
=&(k+1)K(K+1) && \text{(C.83)} \\
\leq&(k+1)2N && \text{(C.84)} \\
\leq&k(k+1)2N, && \text{(C.85)}
\end{aligned}
$$

which implies

$$
\frac{1}{k}(2kN-q_k) \geq \frac{1}{k+1}\left[K(K+1)-q_k\right] \qquad \text{(C.86)}
$$

and hence

$$\frac{(K-1)!}{k}(2kN - q_k) \geq \frac{(K-1)!}{k+1}(K(K+1) - q_k). \tag{C.87}$$

On the other hand, since

$$H(X_{(p_k+1,k+1)}|Z_{[k]}, W_{[p_k]}) \geq H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]}), \tag{C.88}$$

together we have

$$\begin{aligned}
&\frac{(K-1)!}{k}[(2kN - q_k)H(X_{(p_k+1,k+1)}|Z_{[k]}, W_{[p_k]}) \\
&\qquad + q_k H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]})] \\
\geq &\frac{(K-1)!}{k+1}(K(K+1) - q_k))H(X_{(p_k+1,k+1)}|Z_{[k]}, W_{[p_k]}) \\
+ &\frac{(K-1)!}{k+1}K(K+1)(p_{k+1} - p_k - 1)H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]}) \\
+ &\frac{(K-1)!}{k+1}q_{k+1}H(X_{(p_k+2,k+1)}|Z_{[k]}, W_{[p_k+1]})
\end{aligned} \tag{C.89}$$

According to $(C.89)$, now we have

$$
\begin{aligned}
&\frac{(K-1)!}{k}\frac{k}{k+1}[(K(K+1)-q_k)H(Z_{[k+1]},W_{[p_k]})+q_kH(Z_{[k+1]},W_{[p_k+1]})]\\
+&\frac{(K-1)!}{k}[(2kN-q_k)H(X_{(p_k+1,k+1)}|Z_{[k]},W_{[p_k]})\\
&\quad+q_kH(X_{(p_k+2,k+1)}|Z_{[k]},W_{[p_k+1]})]\\
\geq&\frac{(K-1)!}{k}\frac{k}{k+1}[(K(K+1)-q_k)H(Z_{[k+1]},W_{[p_k]})+q_kH(Z_{[k+1]},W_{[p_k+1]})]\\
+&\frac{(K-1)!}{k+1}(K(K+1)-q_k))H(X_{(p_k+1,k+1)}|Z_{[k]},W_{[p_k]})\\
+&\frac{(K-1)!}{k+1}K(K+1)(p_{k+1}-p_k-1)H(X_{(p_k+2,k+1)}|Z_{[k]},W_{[p_k+1]})\\
+&\frac{(K-1)!}{k+1}q_{k+1}H(X_{(p_k+2,k+1)}|Z_{[k]},W_{[p_k+1]}) \qquad \text{(C.90)}\\
\geq&\frac{(K-1)!}{k+1}[(K(K+1)-q_{k+1})H(Z_{[k+1]},W_{[p_{k+1}]})+q_{k+1}H(Z_{[k+1]},W_{[p_{k+1}+1]})]\\
&+\frac{(K-1)!}{k+1}[(2(k+1)N-q_{k+1})H(X_{(p_{k+1}+1,k+2)}|Z_{[k+1]},W_{[p_{k+1}]})\\
&\quad+q_kH(X_{(p_{k+1}+2,k+2)}|Z_{[k+1]},W_{[p_{k+1}+1]})] \qquad \text{(C.91)}
\end{aligned}
$$

$(C.91)$ can be parted into three steps, each is prove in $(C.74)$ to $(C.76)$ respectively. Therefore, we have

$$
\begin{aligned}
&(K+1)!MF+(K-1)!2NRF\\
\geq&\frac{(K-1)!}{k+1}[(K(K+1)-q_{k+1})H(Z_{[k+1]},W_{[p_{k+1}]})+q_{k+1}H(Z_{[k+1]},W_{[p_{k+1}+1]})]\\
&+\frac{(K-1)!}{k+1}[(2(k+1)N-q_{k+1})H(X_{(p_{k+1}+1,k+2)}|Z_{[k+1]},W_{[p_{k+1}]})\\
&\quad+q_kH(X_{(p_{k+1}+2,k+2)}|Z_{[k+1]},W_{[p_{k+1}+1]})]\\
&+k(K-1)!NF, \qquad \text{(C.92)}
\end{aligned}
$$

which fulfill the induction to show Lemma 10. □